DEVELOPING PIXEL-BASED FEATURE SETS FOR INTELLIGENT IDENTIFICATION OF EIMERIA SPECIES FROM MICROSCOPIC IMAGES

Mohamed A Edris Abdalla

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the

Faculty of Engineering and Environment

September 2018

DECLARATION

I declare that no outputs submitted for this degree have been submitted for a research degree of any other institution. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. Any ethical clearance for the research presented in this commentary has been approved. Approval has been sought and granted by the Faculty Ethics Committee.

I declare that the Word Count of this Commentary is 34819 words

Name: Mohamed Abdalla

Signature: Abdalla

Date: 09/07/2018.

Acknowledgements

Before and after anything, I must express my greatest gratitude to almighty Allah who states and urges in the holy Quran that every single human being is required to learn and to spread his/her knowledge. Almighty Allah has been offering me all opportunities and resources to do my best in this research, which I have tried seriously hard to do. Next, I would send my honest appreciation to Dr. Huseyin Seker for his supervision. My gratitude also goes to my father, who passed away a long time ago but his words of encouragement still resonate in my heart. I also send warm thanks to my mother, wife, and children. Finally, I would like to thank all my relatives, friends and colleagues for their essential help and support.

Publications

- Mohamed A. E. Abdalla, Huseyin Seker, "Recognition of protozoan parasites from microscopic images: Eimeria species in chickens and rabbits as a case study", Engineering in Medicine and Biology Society (EMBC) 2017 39th Annual International Conference of the IEEE, pp. 1517-1520, 2017, ISSN 1558-4615.
- Mohamed A. E. Abdalla, Huseyin Seker, Richard Jiang, "Identification of rabbit coccidia by using microscopic images", *Engineering & MIS (ICEMIS) International Conference on*, pp. 1-4, 2016.
- Mohamed A. E. Abdalla, Huseyin Seker, Richard Jiang, "Automated identification of chicken Eimeria species from microscopic images", 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)

ABSTRACT

Morphological features have been investigated in various automation studies to identify different types of medical images. These studies rely on this type of feature, whereas digital images can provide other different features as descriptive characteristics. Pixels can be utilized as feature sets to recognize regions of interests. This research develops pixel values to extract informative features to identify protozoan parasites of the Eimeria genus. Eimeria is a single-celled intestinal parasite which infects humans and animals. Each type of host can be infected with different Eimeria species. Coccidiosis is caused when Eimeria infects animals, which is a rapidly spreading and fatal disease. Its treatment requires the identification of which species has infected the host, but similarities between Eimeria species make identification a very challenging process. Previously, automatic identification was carried out by imitating biological measurements, but these require complex and costly computational processes to extract the desired features. Therefore, this research aims to simplify the feature extraction process considering the use of another type of feature to distinguish between Eimeria species. The features considered do not need complex extraction processes and provide high accurate results. Pixel-based features are analysed by calculating the means of image matrix columns and rows of regions of interests. Features are represented as sets of column features (CF), row features (RF), and combinations of both in (CRF). Moreover, CF, RF, and CRF are extracted from greyscale level and colour images. Therefore, six feature sets are considered, and these are optimized by utilizing five selection and reduction algorithms to minimize the feature space. Furthermore, the extraction of super-pixel feature sets is developed to simplify segmentation. For classification, three classifiers are applied. The 5-fold cross-validation is used to evaluate the results and every experiment is repeated 50 times. Consequently, the results shown are the averages of 50 runs along with values of standard deviation. The proposed method is examined by analysing two microscopic image databases of 4402 images of the 7 Eimeria species in chickens and 2902 images of the 11 Eimeria species in rabbits. The best accuracy results achieved are 96.7% (±0.89%) and 95.85% (±2.4%) for the respective datasets. Finally, the proposed method succeeds in finding simple features to identify Eimeria species, reducing the feature number by 40% of the original size, and super-pixel feature sets are established which give excellent results.

Table of Contents

List of Abbreviations

µm	Micro millimetre
M	The samples in the dataset (the rows of the dataset)
N	The attributes in the dataset (the columns of the dataset)
CF	The column feature set
RF	The row features set
CRF	The column and row feature set
mRMR	Minimum redundancy maximum relevance feature selection
PCA	Principal component analysis
RSV	Remove small value
RFFS	Rectangle frame feature set
PNN	Probabilistic neural networks
F-SDTS	Filtration and steady determination system
SVM	Support vector machines
MTH	Multi-texton histogram
K-NN	K-nearest neighbour classifier
FLD	Fisher linear discriminant
MCSVM	Multi-class SVM
IM-ANFIS	Invariant moments–adaptive network based fuzzy inference system
OPF	Optimum-path forest
LDA	Linear discrimination analysis
CV	Cross-Validation
LOOCV	Leave one out Cross-Validation
ASM	Active shape model
MLP	Multi-layer perceptron
DT	Decision tree classifier
RGB	Red-Green-Blue

MNT	Moore-neighbour tracing
MID	Mutual information difference
MIQ	Mutual information quotient
COV	Covariance matrix
ANN	Artificial neural networks
RBF	Radial basis function
C	Regularization parameter
γ	Kernel width (Gamma)
TP	True positive
SD	Standard deviation
σ	Standard deviation
μ	Overall accuracy rate
NNI	Nearest neighbour interpolation
ACE	E. Acervulina
BRU	E. Brunetti
MAX	E. Maxima
MIT	E. Mitis
NEC	E. Necatrix
PRA	E. Praecox
TEN	E. Tenella
COE	E. Coecicola
EXI	E. Exigua
FLA	E. Flavescens
INT	E. Intestinalis
IRR	E. Irresidua

MAG	E. Magna
MED	E. Media
PER	E. Perforans
PIR	E. Piriformis
STI	E. Stiedai
VEJ	E. Vejdovskyi

List of Tables

List of Figures

CHAPTER 1 **INTRODUCTION**

1.1 **Motivation**

Automated identification studies of Eimeria and other types of parasites have been widely carried out to provide quick, accurate, and reliable approaches to recognition. These studies have exploited image processing and machine learning techniques as a cornerstone in achieving their goals. Currently, imitating the manual steps in the process has been the main strategy used in the majority of published automated methods. For example, biologists in laboratories try to identify parasites under the microscope by measuring the length and width, density, area, symmetry, and texture of cells and objects to identify them from the microscopic images. Similarly, automated techniques have adopted the same diagnostic steps. In other words, computer-based identification techniques apply automated measurements instead of manual ones.

Although some automated studies have achieved very accurate identification rates in different cases, they have encountered numerous difficulties when analysing objects which are highly similar. By and large, they were not able to accomplish very accurate recognition. However, digital images might be analysed in different ways. Basically, digital images consist of huge numbers of values called pixels. These values represent the colours, boundaries, sizes, and texture densities of objects. Normally, the pixels are organized in two dimensional matrices in the case of black and white and greyscale images or in three dimensional matrices for colour images. Apart from parasite identification achieving by measuring the morphological characteristics of objects, this research develops pixel-based feature sets in order to define and extract new easily

extracted features from digital images as well as increasing the accuracy of automated identification methods.

1.2 Scope of Research

Various studies using automated methods have tried to establish the automatic identification of parasites in general and Coccidiosis-caused ones specifically. The aim of those studies was to find precise recognition techniques which provide very accurate levels of discrimination in a short time. They analysed morphological characteristics and classified segmented features using different classification models. However, the extraction of the morphological characteristics involves processes with high computational costs and the resulting accuracy levels have not always been high. Thus, the main target of the present research is to concentrate on developing informative features which will help classifiers obtain higher results to ensure more accurate identification. These new features will not require complex extraction processes because they will be extracted based on the pixel values of the region of interest.

1.3 Aims and Objectives

The aim of this research is to automatically identify Eimeria species from microscopic images based on the analysis of the pixel-based feature sets. The pixel-based features can be very informative descriptors to distinguish Eimeria species. This overall includes pursuing the following five main objectives.

1. Exploring microscopic digital images of Eimeria species to establish the main basis for this research, in order to find good quality image databases. This objective is covered in Chapter 3.
2. Examining the chosen image databases by exploiting previously published automated methods and to determine the accuracy rates they are able to obtain. This objective is covered in Chapter 2 when the literature reviews discussed and three related work have been implemented considering the image databases utilized in this research. The outcomes of those three identification methods are discussed in Chapter 4.
3. Developing pixel-based features and classification models to assess the accuracy rates yielded. This objective is covered in Chapter 3 which explains the proposed methodology for Eimeria identification and the results obtained are reported in shown Chapter 4.
4. Improving the extracted pixel-based features by using feature selection and reduction algorithms. This objective is covered in Chapter 5.
5. Extracting effective super-pixel feature sets to identify the objects considered in every image databases. This objective is covered in Chapter 6.

1.4 Contributions of the Research Study

Three efficient approaches have been proposed in this research to achieve very accurate classification performance based on very reliable features.

1. Chapter 4 considers the segmentation of pixel-based feature sets from greyscale-level as well as colour images.

The extraction of these features has been accomplished after the application of several image processing steps to ensure that all images have exactly the same positions and orientations.

Chapter 4 also discusses the implementation of three different identification methodologies. The aim of utilizing these experiments is to determine the accuracy rates morphological-based methods achieved when exactly the same dataset is analysed.

The classification of the segmented features yields various results, however, the analysis of datasets of colour images always gives the high prediction performance.

2. Chapter 5 proposes improvement in the extracted datasets by applying five methods of feature selection and reduction.

Three of the methods are common used which are ReliefF, mRMR, and PCA. In addition, this research also proposes Remove Small Value (RSV) and Rectangle Frame Feature Set (RFFS) algorithms as new feature selection methods. Utilizing feature selection and reduction algorithms leads to improvements in the resulting classification outcomes with various levels of accuracy.

3. Chapter 6 exploits ReliefF to select super-pixel datasets from both greyscale-level and the colour images.

Depending on the outcomes of the experiments conducted in chapter 5, ReliefF always succeeded in selecting fewer features, which helps the classifiers to yield the best

classification performance. Therefore, ReliefF was applied to select features from image pixels instead of selecting features using CF, RF, or CRF sets.

1.5 Thesis Outline

This thesis is organized in 7 chapters as follows.

Chapter 2 explores the biology and medical aspects of the protozoa parasite of the genus of Eimeria species, including describing infection, lifecycle, diagnosis, and treatment. Furthermore, it discusses automatic recognition studies of parasitic infection which propose computerized identification methods.

Chapter 3 explains the proposed automated identification methodology to classify Eimeria species.

Chapter 4 describes the experiment investigating the segmentation and classification of the pixel-based features adopted.

Chapter 5 discusses the selection and reduction algorithms used to optimize the extracted features.

Chapter 6 utilizes ReliefF feature selection algorithm to select a super-pixel dataset from greyscale level and colour images.

Chapter 7 presents the research conclusions and recommendations for possible future work.

CHAPTER 2 **LITERATURE REVIEW**

This chapter contains two main parts. The first part provides a biological definition and description of the protozoan parasite of the genus Eimeria species, the dangers of Eimeria infection, methods of diagnosis, and treatment procedures. The second part discusses several relevant studies of the automatic identification of different types of internal parasites based on their hosts: humans and various other animals. Previous studies are considered in terms of the methods applied to process, segment, classify, and verify the automatic recognition of regions of interest in images. Furthermore, the classes and sizes of datasets are explored. The presented accuracy rates achieved so far and any limitations in existing research are also discussed.

2.1 **Biological Background**

Parasites are defined as organisms that need hosts to live and reproduce (Mehlhorn, 2012). They acquire food at the expense of their hosts and might infect humans or other animals. Parasites can live inside the body of the host or on the skin. Internal parasites, which live inside the body of hosts, are categorised into two general types: protozoa and helminths. The protozoa are single-celled objects, whilst helminths are different forms of worms. Eimeria, which is also known as Coccidia, is a genus of protozoan parasites. Eimeria species infect animals or humans in their guts leading to damage to the intestines of the host. The disease caused is called Coccidiosis, which is one of the deadliest known parasitic diseases (Kheysin, 2013).

2.1.1 Eimeria Species

(Saif, 2008) states that an animal is susceptible to be infected by several Eimeria species. For example, Eimeria in chickens can be caused by 7 species, there are 11 in rabbits, 22 in dogs, 12 in cows, 8 in pigs, and 7 in turkeys. All of these different species are characterized as having oval shape and their sizes range from 9 to 34 micrometre (µm) in length and 6 to 21 µm in width. Eimeria species for every host have close similarities. Figure 2-1 shows Eimeria species in chickens.

2.1.2 Eimeria Infection

Eimeria infection is called Coccidiosis. It is very dangerous because of the high-speed spread of disease and secondary infection caused. When Eimeria infects a new host, it feeds and multiplies in its intestines and damages the internal intestinal tissue. The infected host will not be able to benefit from feeding due to a poor absorption of nutrients

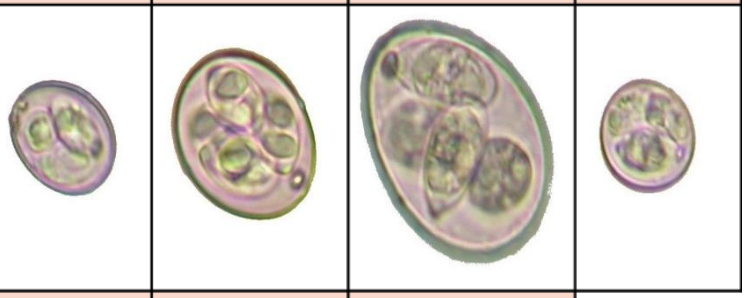
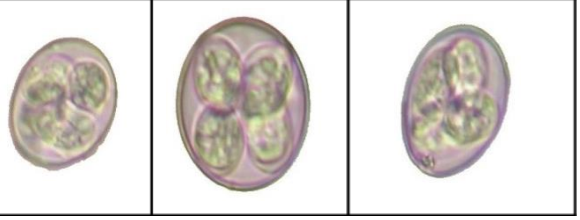
Eimeria spp.	E. Acervulina	E. Brunetti	E. Maxima	E. Mitis
Range of length µm	17.7-20.2	20.7-30.3	21.5-42.5	11.7-18.7
Range of width µm	13.7-16.3	18.1-24.2	16.5-29.8	11.0-18.0
The oocyst under microscope				
Eimeria spp.	E. Necatrix	E. Praecox	E. Tenella	
Range of length µm	13.2-22.7	19.8-24.7	19.5-26.0	
Range of width µm	11.3-18.3	15.7-19.8	16.5-22.8	
The oocyst under microscope				

Figure 2-1 The seven Eimeria species in chickens under the microscope (Saif, 2008)

and also blood loss will occur. Furthermore, bacterial and viral infection will be found as secondary infections because of the intestinal damage. Moreover, Eimeria infection will not stop at the first host. The infected individual will be coccidian infective to others due to the presence of Eimeria oocysts in its droppings (Gardiner, Payer and Dubey, 1988). Coccidiosis infection among animals begins between 6 to 9 days of the first infection, because the Eimeria lifecycle is not identical among different types of hosts. Figure 2-2 depicts the lifecycle of Eimeria in chickens. Simply, the infected animal faeces contain sporulation Eimeria oocysts. As soon as adjacent animals swallow those oocysts, they will start multiplying in the new host's gut and infection will not stop unless effective treatment is applied. In consequence, every infected animal will start dying within 10 – 14 days of infection if not treated. Therefore, Coccidiosis constitutes

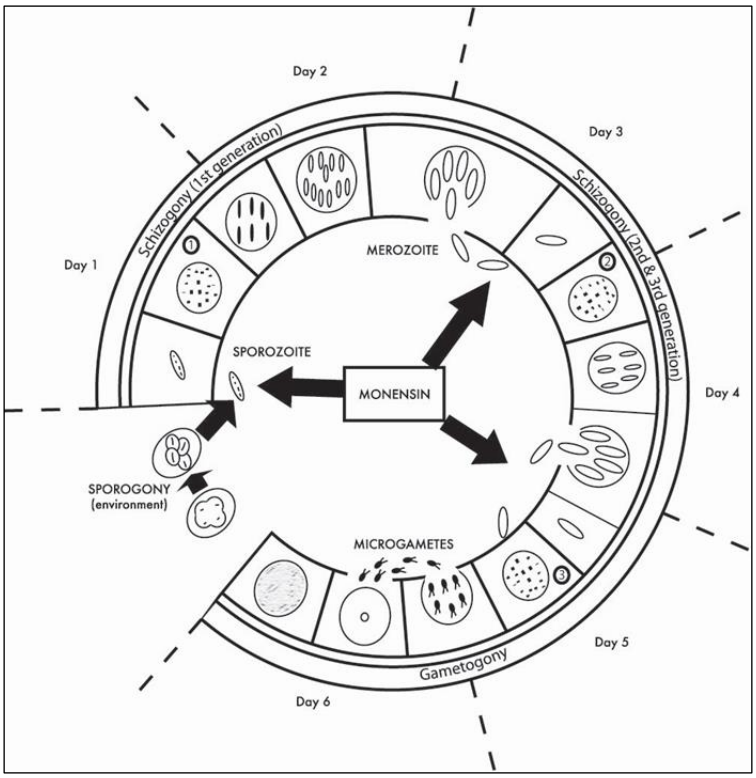


Figure 2-2 Chicken Eimeria lifecycle (Kheysin, 2013)

as an epidemic disease, vets and farm staff are always very keen to detect Coccidiosis in the early stages to avoid any business loss (Gillespie and Bamford, 2013).

2.1.3 **Eimeria Diagnosis**

(Abbas *et al.*, 2011) point out that veterinarians apply diagnostic processes when they notice the symptoms of Coccidiosis. These include loss of appetite, emaciation, weakness, pallor, an unkempt coat of hair, fur, wool or feathers, and severe diarrhoea with faeces containing blood-tinged mucous or blood. The diagnosis of Eimeria infection can be concluded in three different ways. The easiest and most common is to examine faecal samples under the microscope. The oocysts of Eimeria species are very noticeable on microscopic slides. However, the identification of Eimeria species is very difficult because of the strong resemblances between different species. Therefore, vets might apply another method. The dissection of infected animals is an alternative (Gardiner, Payer and Dubey, 1988) which is more accurate than examining samples under the microscope, because every Eimeria species tends to accumulate and cause damage in a particular area of the intestine. However, the autopsy has to be performed with one hour after animal death, otherwise identification will not be effective. Figure 2-3 shows the turkey intestines infected by seven Eimeria species (*E. Adenoeides, E. Dispersa, E. Gallopavonis, E. Innocua, E. Meleagridis, E. Meleagrimitis, E. Adenoeides* and *E. subrotunda*). It shows that every Eimeria species gather and cause damage in different place of the gut of the host. The third method of identification is to apply DNA analysis to distinguish between Eimeria oocysts (Zonaed Siddiki *et al.*, 2014). This is

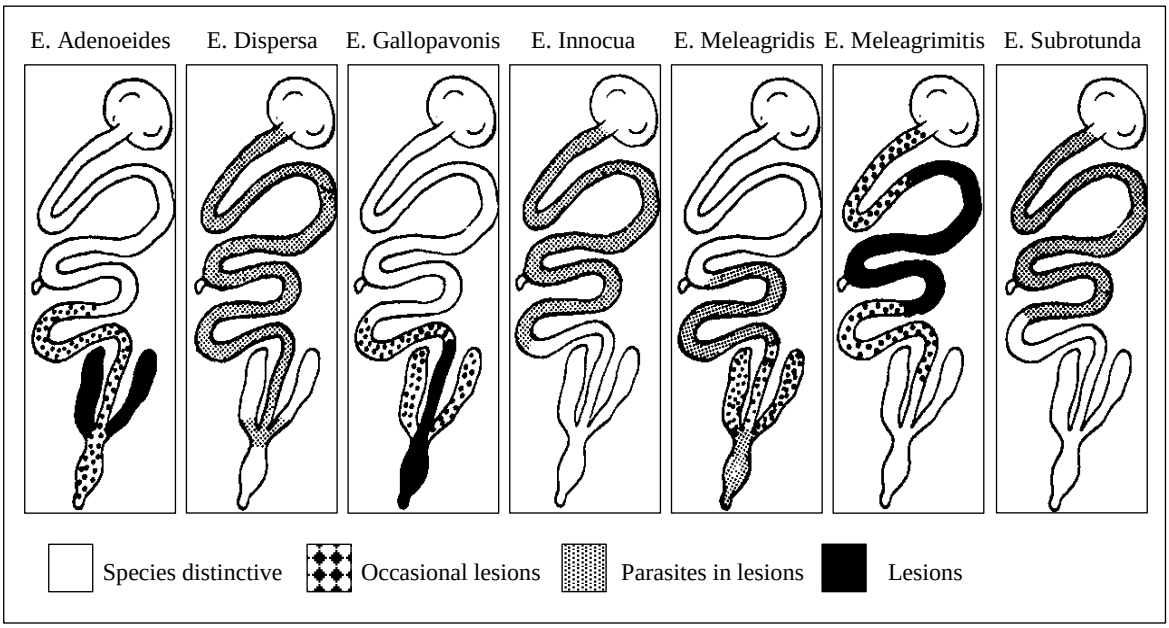


Figure 2-3 dissection of infected turkey intestines by Eimeria species (Becker, 1976)

the most accurate identification method, but it is a complex and lengthy process. Taking a long time to identify species is not helpful in dealing with Coccidiosis.

2.1.4 **Eimeria Treatment**

(Abbas *et al.*, 2011) explain that the correct anti-coccidiosis treatment requires precise identification in order to deliver effective drugs during the early stages of infection. Otherwise, various types of anti-coccidiosis drug must be prescribed to stop the disease spreading as well as the treatment of secondary infections. Every animal will be involved in these treatments which will be very expensive. Finally, early correct coccidiosis diagnosis and therapy are always highly recommended because severe infection will prevent animals from efficiently recovering to their market weights.

2.2 Computerized Identification of Parasites

Since the last century, researchers have sought to provide methods for the reliable automated identification of Eimeria species and other internal parasites in human and animals based on the analysis of digital optical microscopic images. Studies of automated recognition of parasitic infection in both humans and animals are discussed below.

2.2.1 Automatic Detection and Identification of Parasites in Humans

There are several automatic recognition studies which have presented different ways to diagnose and identify many parasite infections. In general, the differences between those studies are in the type of parasite concerned, the distinctive segmented characteristic considered, and the classification processes used. Those studies are discussed below and Table 2-1 shows a summary.

Nine types of parasite cysts in human from a database of 450 microscopic images were analysed and automatically classified by (Tchinda *et al.*, 2015). They segmented features based on pixel values. Cyst images were first prepared in 12×12 image matrices, so, giving a feature space of 144 features. Before the classification stage, the feature vector was reduced to 2 features using principal component analysis (PCA) algorithm. For the classification phase, probabilistic neural networks (PNN) was used to classify the human protozoan parasite images. PNN is a type of feedforward neural networks (Suzuki, 2011), and then the dataset was randomly divided into 50% training set and the rest a testing set, accuracy rate of 100% was obtained. However, this result could not be considered reliable as they did not repeat the experiment many times with random changes to the training and testing sets, because testing different samples might affect the outcomes.

(N.Suzuki *et al.*, 2013) examined the five morphological features area, length, width, boundary length, and roundness to identify the two human parasites roundworms and whipworms. They analysed 100 microscopic images of each parasite. For classification, they presented an if-condition-based classification simple method called the Filtration and Steady Determination System (F-SDTS). An overall accuracy rate of 93.5% was achieved. Intensive image processing and segmentation procedures were used to enhance the detection of candidate objects. However, F-SDTS model was simply based on if-statement. Therefore, testing its effectiveness requires the use of different image databases, due to the fact that the two parasites considered had very differentiated characteristics (Mehlhorn, 2012).

(Saito *et al.*, 2015) applied support vector machines (SVM) and optimum-path forest (OPF) classifiers to identify 15 common human intestinal parasites. Using k-mean technique they clustered three different databases of microscopic images: d1:1455, d2:4458, and d3:141059. For feature extraction, three groups of features were segmented from clusters based on colour, shape, and texture. The evaluation stage randomly divided images into 80% for training and 20% as testing sets. The overall final results of the implementation of the classification method repeated 10 times with random selection of the training and testing sets were 96.99%, 91.58%, and 88% accuracy rates for d1, d2, and d3 respectively. Achieving these significant results with different datasets shows that it was very a reliable recognition technique. However, the number of clusters used to achieve these accuracy results was not mentioned.

(Flores-Quispe *et al.*, 2014) utilized multi-texton histogram (MTH) descriptors, which is a technique used to extract colour image microstructures according to relationships between adjacent pixels and the detection of edges. MTH methods extracted 82 features

from a database of 2053 microscopic images of 8 types of parasitic eggs in humans. Those extracted features represented the image colour and texture characteristics. Classification of those eggs was applied using SVM. The dataset was randomly divided into 50% training and 50% testing sets. The study achieved a correct classification rate of 92.16%.

(Tek_B, Dempster and Kale, 2010) investigated microscopic images of blood samples to detect and classify candidate objects in one of the four malaria species of the genus Plasmodium (*Falciparum*, *Vivax, Ovale*, and *Malariae*) and white blood cells, platelets, or artefacts. The dataset included 4100 images. From every image, 83 features were analysed based on the image histogram, area, and shape measurement (Tek, 2007). The training and testing sets were built by the leave-one-out cross validation method. Then, three classifiers were used to examine the features: K-NN, back propagation neural networks (BPNN), and the Fisher linear discriminant (FLD). The best accuracy rate was 93.3% obtained by K-NN.

(Avci and Varol, 2009) examined the Hu's invariant moments as a feature set from a database of 960 images to recognize 16 human parasite eggs. Half of the dataset was randomly chosen as a training set. Multi-class SVM classifier (MCSVM) was applied to the training set. An overall correct predication rate of 97.7%. was obtained. Moreover, 100% of correct classification rate was found from four of the observed classes.

(Dogantekin *et al.*, 2008) also analysed the same 7 features of Hu's invariant moments feature to identify another 16 human intestinal parasites from a balanced dataset of 2400 microscopic images. The Hu's invariant moment process requires rotation of images at different scales to extract features. They rotated images at steps of 15^{o} from 0 to 165^{o} and extracted 7 features from every rotated image. So, every image generated 7×15

features. In the classification stage, an invariant moments–adaptive network based fuzzy inference system (IM-ANFIS) was proposed to classify the parasite eggs. However, the stage of training and testing was not very reliable, because the classifier was trained by the features of all rotated images in steps of 0^o and 60^o, then tested by the features of the rest rotating angles. That might allow the classifier to cheat resulting inaccurate performance rates that reached 95%. Therefore, examining classifier performance with different methods of dataset division is required especially with a large pool of evaluation methods available.

Other studies have analysed microscopic images to count parasite species by determining whether objects are parasite species or artefacts without any identification of parasites. (Arco *et al.*, 2015) investigated images of human microscopic blood slides to determine objects if were Malaria parasite cells or white blood cells. They examined a database of 475 microscopic images, and implemented segmentation by exploiting morphological operations. Their aim was to count the number of malaria cells in every image and discard white blood cells. That study achieved a correct enumeration rate of 96.46% of malaria species with no identification. This high result was obtained from use of the pixel values to detect if the objects were malaria cells or white blood cells. (Peixinho *et al.*, 2016) utilized deep learning techniques to diagnose and count 15 common human intestinal parasites. They investigated 12 descriptive features: the colour histogram, object area, texture, perimeter, symmetry, major and minor axes, the best-fit ellipse, object energy, entropy, and variance, and the homogeneity of the co-occurrence matrix. The dataset size contained 16,437 objects which were divided into 5112 protozoa, 9568 helminth, and 1757 larvae. They achieved high rates of accurate detection and counting of detecting and counting of parasite objects up to 96.79%.

Table 2-1 Summary of the studies of the automated identification of human parasites

Candidate objects	Object number	Dataset size	Balanced data	Segmented features	Dataset division	Classification model	Comments
protozoan parasites (Tchinda *et al.*, 2015)	9	450	Unknown	144 features of pixel values from 12×12 image sizes	50% training set, 50% testing set	Probabilistic neural networks (PNN)	**Pros:** - The features considered were easily extracted - PCA succeeded in sharply reducing the feature space from 144 to 2 features. - Obtained 100% accuracy rate **Cons:** - Dataset division was not effective, because half of dataset was not examined in the testing task
Intestinal parasites (Saito *et al.*, 2015)	15	d1: 1455 d2: 4458 d3: 141059 images	Unknown	3 feature groups based on colour, shape, and texture	80% training set, 20% testing set	SVM And OPF	**Pros:** - Examined three datasets and Obtained good results from all datasets.

Candidate objects	Object number	Dataset size	Balanced data	Segmented features	Dataset division	Classification model	Comments
Intestinal worms (N.Suzuki *et al.*, 2013)	2	200	Yes	5 morphological features	Unknown	Filtration and steady determination system (F-SDTS)	**Pros:** - Exploring many image processing and segmentation techniques to enhance detection. - 93.5% accuracy rate. **Cons:** - Method of division of dataset to training and testing sets was not declared. - The difference between the objects analysed was very clear. The first class is long, but the second has a circular shape. - The classification process was built on if-condition statement.
Parasite eggs (Flores-Quispe *et al.*, 2014)	8	2053	No	82 texture and colour features	50% training set 50% testing set	SVM	**Pros:** - Intensive image analysis. **Cons:** - The class with the highest accuracy rate had the highest sample size in the dataset. However, classification of those classes with smaller samples provided lower accuracy rates. - They did not utilize feature selection algorithms which might find less attribute number with higher rates.

Candidate objects	Object number	Dataset size	Balanced data	Segmented features	Dataset division	Classification model	Comments
Malaria species (Tek_B, Dempster and Kale, 2010)	7	4100	Unknown	83 features	Leave-one-out	K-NN FLD BPNN	**Pros:** -Three classifiers were used to examine the extracted features, and all obtained good results. **Cons:** -No any attempt to reduce the feature space using feature selection and reduction algorithms
Intestinal parasite egg (Avci and Varol, 2009)	16	960	Yes	7 Hu's invariant moments	50% training set, 50% testing set	MCSVM	**Pro:** -Analysis of balanced dataset makes the classification outcomes more reliable. **Cons:** -Some samples might not be examined in both training and testing sets. That may reduce the reliability of the final results.
Intestinal Parasite egg (Dogantekin *et al.*, 2008)	16	2400	Yes	7 Hu's invariant moments	All images in training stage	IM-ANFIS	**Pro:** -Balanced dataset **Cons:** -Using the images in the training set phase could lead to imprecise results.

2.2.2 Automatic Detection and Identification of Parasites in Animals

Machine leaning techniques based on analysis using image processing methods have also been proposed in numerous studies of automated diagnosis and identification of various internal parasite species. These studies provided different levels of accuracy via the analysing of several extracted descriptors. Table 2-2 summarizes those studies and the sections below discuss them based on the hosts concerned.

2.2.2.1 Bovine

(Sommer, 1998a) researched the identification of 9 Eimeria species in cattle from a database of 810 digital microscopic images. The segmented characteristics were morphological aspects based on clustering processes. For classification, an agglomerative clustering technique was applied to distinguish between the species analysed. Classification based on the clustering technique reached an overall correct prediction rate of 72%.

(Sommer, 1996) investigated the recognition of five bovine nematode parasite eggs via analysis of 24 morphological features of 4207 eggs from microscopic images. The use of linear discrimination analysis (LDA) provided classification performance percentage rates of 76.3%, 90.8%, 87.8%, 91.1%, and 83.8% for the eggs of *Ostertagia ostertagi*, *Cooperia oncophora*, *Haemonchus placei*, *Trichostrongylus axei*, and *Oesophagostomum radiatum* respectively. The overall average accuracy for these species was 85.8%.

(Sommer, 1998b) then analysed textural features only from 111 microscopic images of three intestinal parasite eggs in cattle: *Ostertagia ostertagi, Cooperia oncophora,* and *Oesophagostomum radiatum*. Leave-one-out cross validation technique was

adopted to evaluate prediction performance. An overall accuracy rate of 91.2% was gained when the LDA classifier was used to process those features.

2.2.2.2 Chickens

(Kucera and Reznicky, 1991) proposed one of the earliest semi-automated recognition methods in this field. They considered the seven Eimeria species by analysing oocyst length, width, and index shape. Theirs was quite a simple identification technique, because detecting the oocyst edge was performed manually. Analysis of these features only, given that species analysed had extreme overlaps in their features, did not help in obtaining high results. The error rate reached 44% in some cases.

(Castañón *et al.*, 2007) analysed 13 features of three characteristic categories: curvature (mean of curvature, curvature standard deviation, entropy curvature), size and symmetry (major and minor axes, symmetry through major and minor axes, area, entropy of oocyst content), and the internal object structure (angular second moment, contrast, inverse difference moment, the texture entropy). These features were segmented from an unbalanced database of 3891 microscopic image. The study applied Bayesian classifier to identify these samples. Furthermore, the experiment was repeated for 9 times to take account of the method of dataset division from 10:90 to 90:10 of training and testing sets. The overall accuracy rate was 85.75% which represented the average of all of the experiments. In an attempt to increase performance, they tried to pick equal numbers of observations from the dataset to avoid classification bias. Therefore, they randomly chose 320 samples from every class and then classified them using the same steps. However, this classification did not actually provide better results. Regarding the importance of the features analysed, the geometry and texture groups showed the highest significant contributions to

increasing classification performance. Although the curvature group had a low role in affecting the accuracy rate, the attribute of curvature standard deviation even less effective as it reduced performance by 0.15%.

Although considering 13 diverse features yielded 85%, classification of the same features obtained lower level of accuracy when the publisher analysed Eimeria species in rabbits. However, this research will examine those features to find out how accuracy rate can they provide.

According to the limitations faced when morphological characteristics were analysed to differentiate between Eimeria species in various hosts, chickens in particular, a large number of studies have considered the analysis of DNA as an alternative descriptor to provide precise identification (Oliveira *et al.*, 2011), (Raj *et al.*, 2013), and (Brown Jordan *et al.*, 2018).

2.2.2.3 Porcine

(Plitt *et al.*, 1999) proposed a semi-automated classification method to identify five Eimeria species by considering three morphological descriptors of colour, length, and width. A database of 175 images were considered to analyse and segment the features. Then the three features segmented were processed using decision tree classifier. Individually, 100% was achieved for one of the species investigated. A correct classification rate, whilst the overall accuracy rate was 97%.

(Daugschies, Imarom and Bollwahn, 1999) utilized the decision tree classifier to identify seven species of porcine Eimeria. The features segmented were oocyst length, width, and colour. With a sample of 4088 oocyst images, the overall classification performance was 93%.

Moreover, identification of Porcine Eimeria species based on the analysis of DNA has been widely applied to avoid any mis-classification challenges (Wiedmer *et al.*, 2017) .

2.2.2.4 Fish

(Han *et al.*, 2011) applied two automated stages to diagnose and identify infections of olive flounders from microscopic images. Besides exploring three parasites species: *Scuticociliatosis, trichodinosis,* and *white spot*, they tried to identify 5 bacterial and 6 viral objects. Therefore, in the first stage the candidate objects were classified into parasites, bacteria, or viruses. The second stage was to identify species of the object to its particular. The two features analysed were morphological characteristics. The database included 60 images of the three parasite species. It was divided into 50% for training and the rest for testing sets. Classification was performed according to the correlation coefficient. The performance results were not shown, but seemed to be low. Low accuracy rates were expected since only two morphological features were considered, whereas more were needed because the shapes of the two species overlap. The authors also stated that improving classification performance in the second stage would require the exploration in the feature extraction phase to find more informative features.

(Park, Oh and Han, 2007) applied quite similar steps to those in the previous study which was conducted later. They examined exactly the same morphological characteristics of the same parasite species using different database of 80 digital microscopic images. Also, the classification model was the same. However, PCA was used to reduce the feature space before the classification stage. The evaluation of the prediction model performance was based on dividing the dataset equally into two halves for training and testing sets. An accuracy rate of 87% was achieved.

Furthermore, the authors tried to divide the features into two specific categories of geometric and the polar coordinate system in order to determine which had the great contribution. Although the extraction process of geometric features was more complicated, their contribution was lower than that of the polar coordinates. The classifier achieved 80% from polar features but from geometrics, it obtained less than 75%.

(Kalafi *et al.*, 2016) examined 4 species of Monogeneans: *Sinodiplectanotrema malayanus, Trianchoratus pahangensis, Metahaliotrema mizellei* and *Metahaliotrema*. The features utilized were 10 morphological descriptors: Euler number, perimeter, area, area density, perimeter density, centre of bounding box, length of bounding box, width of bounding box and orientation of bounding box. This number of features was reduced using Linear Discriminant Analysis (LDA). The optimum feature vector in this case contained 3 features. For classification, K-NN classifier was exploited with K=10. Evaluation of classification performance was achieved by dividing the dataset in three ways. The division used 50% of the images as training and 50% as the testing set. The second utilized Leave-one-out cross validation methods. The third way was 10-fold cross validation. Overall accuracy rates obtained were 90%, 92.25%, and 92.5% respectively based on the dataset division methods used.

(Ali, Hussain and Man, 2015) analysed different types of features to identify 4 species of Gyrodactylus: (*G. colemen*, *G. derjavinoides*, *G. salaris*, and *G. truttae*) from database of 78 digital microscopic images. Extraction of those different features were based on the active shape model (ASM) technique. ASM detects a number of landmark points for the region of interest. They selected N=110 points. Then PCA was applied to reduce the number of points. They found that the optimal number of landmark points was N=48. Four common prediction models were used: K-NN, LDA, MLP, and SVM.

All classifiers were trained with 9-fold and then tested with 1-fold of cross validation algorithm. As a result of repeating the experiment for 10 runs using different testing fold every experiment, this study provided 98.72% of the best overall performance of accuracy rate using Multi-layer perceptron (MLP).

The study of utilizing ASM to identify the parasites showed that another type of features can be very informative descriptors to distinguish the parasites. Therefore, as a part of this research, the published method of ASM will be implemented on the image databases of Eimeria species to find out how accuracy rate can it yield.

Table 2-2 Summary of the studies of the automated identification of different animal parasites

Candidate objects	Object number	Dataset size	Balanced data	Segmented features	Dataset division	Classification model	Comments
Bovine egg parasite (Sommer, 1996)	5	4207	No	24 morphological features	Leave-one-out	LDA	**Pros:** -One of the earliest automated studies in this field. **Cons:** -The dataset was unbalanced. -The role of features was only significant with the first 10 features. The rest of features did improve the classification performance.
Chicken Eimeria species (Castañón *et al.*, 2007)	7	3891	No	13 morphological features	10:90 to 90:10 of training and testing sets	Bayesian	**Pros:** -Provided a public image dataset. -Utilized varied morphological features. **Cons:** -Some features let the classifier confused.
Porcine Eimeria species (Daugschies, Imarom and Bollwahn, 1999)	7	4088	Unknown	3 morphological features	Not shown	DT	**Pros:** - One of the earliest studies in this field which obtained excellent classification performance. **Cons:** -Training and testing sets were not properly explained. -Number of extracted features was too low to build a reliable recognition technique.

Candidate objects	Object number	Dataset size	Balanced data	Segmented features	Dataset division	Classification model	Comments
Fish parasite species (Park, Oh and Han, 2007)	4	80	No	8 morphological features	50% training set 50% testing set	Correlation coefficient	**Pros:** -Detailed research exploring the importance of every group of the features. **Cons:** -The classification technique was very simple. -Dataset size was quite small.
Fish parasite species (Kalafi *et al.*, 2016)	4	102	Yes	10 morphological features	10-fold CV, Leave-one-out, and 50% training set 50% testing set	K-NN	**Pros:** -Evaluation of classification performance examined 3 different methods to build training and testing sets. -Extracted feature space was reduced by LDA selection method. **Cons:** -Results obtained from 80 images only. -Any image, which contained noise, was intentionally discarded. Validation cannot be accurate when using only perfect images.

Candidate objects	Object number	Dataset size	Balanced data	Segmented features	Dataset division	Classification model	Comments
Fish parasite species (Ali, Hussain and Man, 2015)	4	78	No	110 features based on ASM	10-fold CV	LDA, K-NN, MLP, and SVM	**Pros:** - Use of ASM in this the field of parasite recognition. - Exploring four classifiers to examine the segmented features. - Reduction of the dataset size using PCA from 110 to 42 features. **Cons:** - The selection of 110 of ASM landmark points was not justified.

2.3 Discussion of the Methodologies Reviewed

Studies of automatic identification using image processing need to include both image processing and segmentation phases, followed by classification and evaluation phases. As shown in this chapter, the first two phases have been widely researched in very various ways. Considerable numbers of feature types have been utilized to find accurate identification techniques. Image processing cannot be unified because each sort of image has different conditions of capturing (Acharya and Ray, 2005). Therefore, researchers are required to find the best processing techniques for their images of interest. Regarding the segmentation stage, the use of the morphological characteristics has dominated. The extraction of this type of features requires very complex and costly computational processes (Rege *et al.*, 2013). However, these related studies do not agree about which type of the morphological features may be more reliable. For example, (Castañón *et al.*, 2007) figured out that the curvature characteristics do not improve the accuracy of outcomes in the case of chicken parasite images, on the contrary, one feature of the curvature characteristics caused decreasing the accuracy results. Meanwhile, (Park, Oh and Han, 2007) declared that geometric features do not provide high contribution. In practice, a huge number of computerized studies have simply automated manual identification processes. In other words, the automatic identification of parasite species exactly follows the processes of identification used in biological laboratories. Biologists always measure the sizes, circularity, and of texture of object structures to identify parasite species under the microscope (Becker, 1976) and (Mehlhorn, 2012) . Apart from extracting all of those morphological features, which often requires highly complex processes, utilizing the

pixels in digital images as descriptive features to identify parasite species from microscopic images can be very informative. (Suzuki *et al.*, 2013) mentioned that the values of the red and green image channels are more significant than the blue channel when they use as descriptors. (Tek, Dempster and Kale, 2009) pointed out that colours simplify object detection and identification in computer vision techniques, and colours can be considered either in terms of pixels or based on histogram curves. Furthermore, (Abdulla, Jayesh and Nair, 2015) presented an automated identification method based on colour-pixel features to classify microorganisms from urine microscope images. (Saito *et al.*, 2015) utilized the colour pixel values to cluster candidate objects in microscopic images. (Cong *et al.*, 2015) investigated image colours to segment diseased areas. They replaced patch-based extraction with super pixel-based feature extraction, focusing on the degrees of colour in microscopic images. Their method with colour-based super-pixels yielded results higher than those from the patch technique. Therefore, it could be concluded that, colours in digital images are very important factors. They need to be taken into account in image processing stages such as when separating the background from the foreground and detecting object boundaries. In addition, the segmentation stage might rely on image colour characteristics as significant features. Moreover, the extraction of image colours does not require a costly processing. Colour information simply can be collected from the pixel values in the case of colour images or after normalizing the colour images to greyscale level, which is also a simple process.

The second aspect of automatic identification is to find a compatible classifier models to use in processing the features extracted. As shown in this chapter, different classifiers have been applied and various levels of accuracy have been yielded according to the method used to dividing the dataset into training and testing sets.

Therefore, this research considers more than one classification model to process the descriptors extracted.

In conclusion, the related work discussed in this chapter, which utilized various morphological descriptors, did not show that a particular group of those morphological characteristics can be informative features in order to differentiate between the regions of the interest. Moreover, in most cases those features require very complex computational processes for extraction. Therefore, the hypothesis of this research is to find another type of features which are described as simple extracted and very informative features.

2.4 **Summary**

This chapter has described many studies of automatic identification which classify the candidate object depending on diverse features. Those features always represent different aspects of the morphological characteristics. However, colours might be very significant features. Furthermore, colours do not require high-cost computational processes. Thus, analysis of colour in microscopic image can be extracted based simply on the pixel values. This analysis is considered in this thesis in more than one form such as greyscale level, colour, and the analysis of RGB channels in an image individually. This research concentrates on finding robust and high accuracy rates from the extraction of simple pixel-based feature sets.

CHAPTER 3 METHODOLOGY AND DATASETS

This chapter explains the stages of the proposed methodology. This methodology aims to analyse the microscopic images and to present a precise automatic classification with the minimum possible error rates. These stages start with image processing followed by extracting the desired pixel-based features. The extraction stage includes different ways to explore the pixel-based feature datasets. Afterwards, a discussion is provided of the three classification models which will be used to identify Eimeria species based on the segmented features. Finally, the techniques used to evaluate classification results are explored. This chapter also describes the databases of images utilized in this research. The stages of the proposed methodology are discussed below, and its framework is depicted in Figure 3-1.

3.1 Image Processing

The image processing steps used vary between studies. In general, image processing aims to make images ready for the extraction of informative features (Costa and Cesar-Jr., 2001; Acharya and Ray, 2005). This research requires the precise detection of the Eimeria oocyst in every image. The following processing steps are applied to detect regions of interest.

3.1.1 Greyscale Level Conversion

Images are captured as colour photos. Therefore, they will be converted to be in greyscale levels. Every colour image I(R,G,B) is converted to in greyscale level image $i(x,y)$ as follows:

$$\boldsymbol{i(x,y) = \alpha R + \beta G + \gamma B} \quad 3.1$$

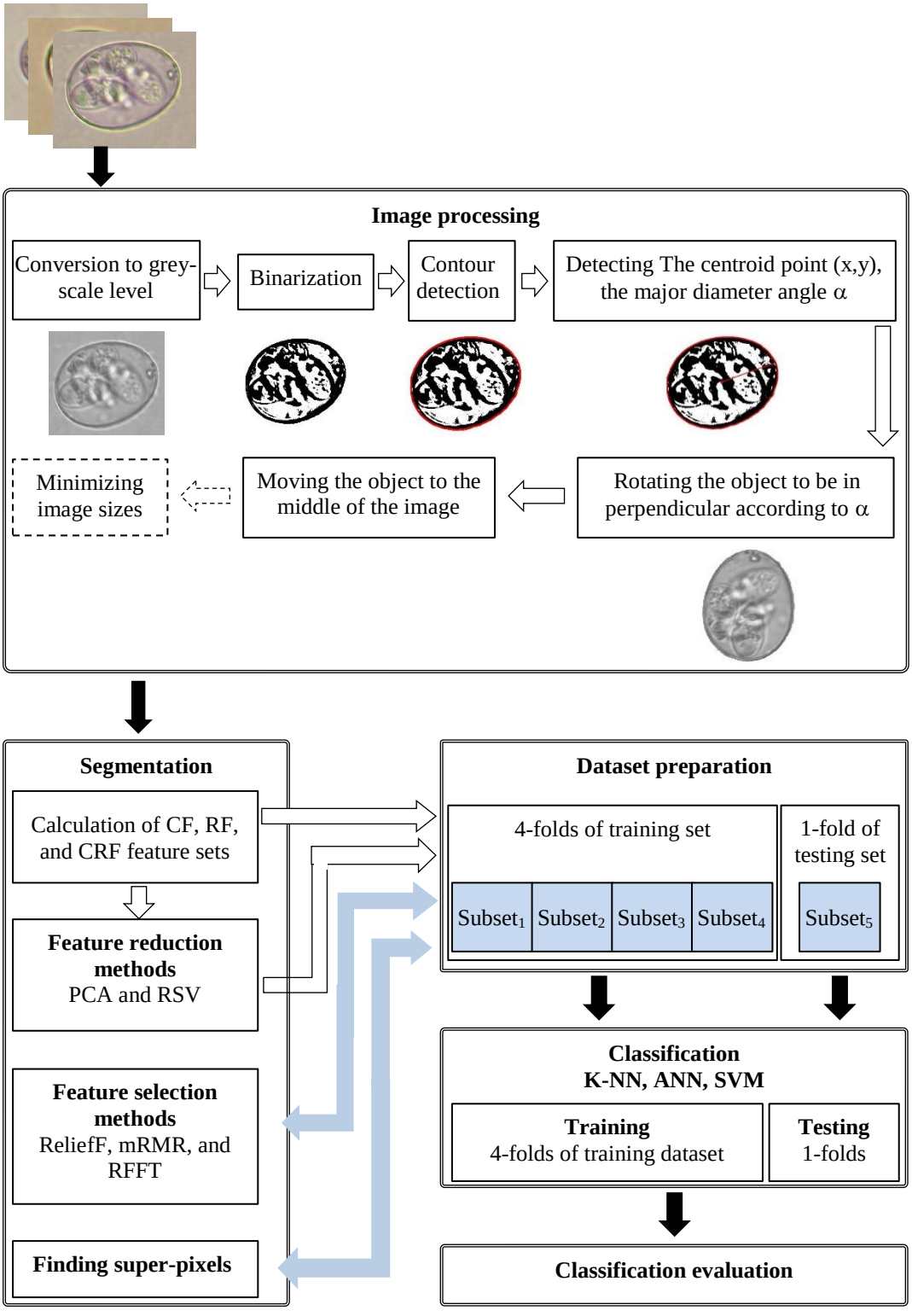


Figure 3-1 The framework of the proposed identification methodology

where *i(x,y)* is the new image of greyscale level, α, β, and γ are constants which multiply in every pixel value of the three channels R, G, and B of the colour images respectively. There are various values of these three constants (Kanan and Cottrell, 2012), and no obvious variance with different values of the three constants. The proposed methodology obtained the best results when 0.3, 0.59, and 0.11 were assigned to α ,β , and γ respectively.

3.1.2 **Image Binarization**

Image binarization has been achieved by calculating the threshold value T automatically from every greyscale image I(x,y). There are several techniques to calculate the threshold value such as Otsu's, Niblack, Local-variance-based methods (Zheng, Ye and Tang, 2017). The best detection results have obtained using Otsu's method (Otsu, 1979; Moallem and Razmjooy, 2012). The value of T has then been exploited to discriminate the image foreground from background to create a binary image f(x,y):

$$\boldsymbol{f(x,y)} = \begin{cases} \boldsymbol{1\ if\ i(x,y) \geq T} \\ \boldsymbol{0\ if\ i(x,y) < T} \end{cases} \qquad 3.2$$

3.1.3 **Detection of Oocyst Contour**

Detecting object edges from binary images is achieved by applying the Moore-Neighbour tracing (MNT) algorithm (Reddy, Amarnadh and Bhaskar, 2012; Seo *et al.*, 2016). The MNT algorithm aims to detect on exterior boundary by finding the first black pixel value where the image background colour is white while the forecolour/object colour is black, and then it searches for adjacent black pixels from its eight neighbouring pixels until arriving back to the first pixel position, as illustrated in Figure 3-2.

Begin

Set *BorderPoints* to be empty.

- From bottom to top and left to right scan the cells of *i* until a black pixel s is found.
- Insert s in *BorderPoints*.
- Set the current pixel *p* to be the starting pixel s.
- Turn with the clock i.e. visit the adjacent pixels of *p*.
- Update *p* i.e. set it to be the current pixel.
- While *p* not equal to s do

If the current pixel *p* is black

Insert *p* in *BorderPoints* and turn with the clock
(visit the adjacent pixel of p).
Update *p* i.e. set it to be the current pixel.

Else

turn with the clock (visit another adjacent pixel of p).
Update p i.e. set it to be the current pixel.

End while

End

Figure 3-2 The steps of MNT to detect the boundary of the object of interest

where; *i* is a binary image, *BorderPoints* is the array of the border, s is the first black pixel, and *p* the current pixel.

In some cases, the starting black pixel belongs to noise not to the region of interest. Therefore, the proposed methodology will repeat the detection until detecting every object in the image and the region of interest will be the biggest object.

3.1.4 Determining Centroid Point of the Object

For any contour vector determined using the MNT algorithm which represents discrete data, the centroid point coordinates *Centroid*(x, y) can be found using the following equation:

$$Centroid(\mathrm{x},\mathrm{y}) = \frac{\mathbf{1}}{\boldsymbol{points}} \times \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{x_i}, \boldsymbol{y_i}) \qquad 3.3$$

where; *points* is the number of contour points of the object, and (x_i, y_i) are the coordinates of the contour points.

3.1.5 Finding the Major Radius

The shape of Eimeria oocysts is oval, they are not regular circles. Therefore, the major axis is the longest straight line between two points on the oocyst boundary which crosses the centroid point. That vector must have a straight-line equation (Costa and Cesar-Jr., 2001).

$$\boldsymbol{y = slope.x + b} \tag{3.4}$$

where; x, y are the coordinates of the point, slope is the line slope, b is a constant.

3.1.6 Rotating Region of Interest

The appropriate rotated angle θ which makes an oocyst in perpendicular to the horizontal axis based on its major radius, can be determined based on the slope of the radius, as shown in equation 3.5:

$$\boldsymbol{slope = \frac{y_2 - y_1}{x_2 - x_1} \rightarrow slope = \tan(\theta) \rightarrow \theta = tan^{-1}(m)} \tag{3.5}$$

where; (x_1,y_1) and (x_2,y_2) are the two coordinate points on the radius line.

3.1.7 Resizing Image Dimensions

There are several categories of resizing algorithms. These algorithms are able to minimize the dimensions of image matrices; which in turn leads to reducing the space of feature set. This research examines images using three types of algorithms of the non-adaptive image interpolation: Nearest Neighbourhood, Bilinear, and Bicubic algorithms. The nonadaptive interpolation techniques are simpler than the adaptive algorithms and because they work on the pixel domain. So, they are quicker and uncomplicated which leads to short run time. (Prajapati, Naik and Mehta, 2012; Mahajan and Harpale, 2015). In this research, image dimensions will be reduced by various degrees. The aim of reducing image sizes is to investigate how pixel-based features will be affected by smaller image dimensions.

3.2 Feature Extraction

Feature extraction aims to represent the candidate objects with informative descriptors (Acharya and Ray, 2005). This research examines pixel-based features of the processed images to identify candidate objects. Fundamentally, the features are extracted by calculating the mean of image pixel values. Three feature sets will be segmented from every image I. The grey images are represented in matrices of two dimensions whilst the colours are three dimensions.

3.2.1 Column Features (CF)

Column features constitutes the mean of pixel values from every column of the oocyst image. Its size will be equal to the image column number ColNo. For colour images, the space of features will be the column number multiplied in 3.

- Grey CF: every element cf will be calculated from all of the pixels P of every column, as illustrated below:

$$\textbf{Grey } \boldsymbol{CF} = [\boldsymbol{cf_1}, \boldsymbol{cf_2}, \ \dots, \boldsymbol{cf_{ColNo}}] \tag{3.6}$$

$$\boldsymbol{cf_z} = \frac{\mathbf{1}}{\boldsymbol{RowNo}} * \sum_{\boldsymbol{i=1}}^{\boldsymbol{RowNo}} \boldsymbol{P_{(i,z)}} \tag{3.7}$$

where *z* is the column number (*z*=1, 2, 3, … ColNo), RowNo is the number of rows of the image matrix.

- RGB CF: the features of this dataset are extracted using the following equations:

$$\boldsymbol{RGB\ CF} = [\boldsymbol{cf_{1RGB}}, \boldsymbol{cf_{2RGB}}, \boldsymbol{cf_{3RGB}}, \dots, \boldsymbol{cf_{ColNo\ RGB}}] \tag{3.8}$$

$$cf_{z\,RGB} = \frac{1}{RowNo} * \sum_{i=1}^{RowNo} \left[P_{i,z(R)}, P_{i,z(G)}, P_{i,z(B)}\right] \quad 3.9$$

where R, G, and B are the channels of the colour image.

3.2.2 Row Features (RF)

Row features (RF) consist of the extracted features based on the pixels of image rows RowNo. As with the CF groups, RF will be considered from greyscale and colour oocyst images.

- Grey RF: the features of this dataset are acquired depending on the pixels *P* of image row number RowNo as shown below:

$$\mathbf{Grey\ RF} = [rf_1, rf_2, \ldots, rf_{RowNo}] \quad 3.10$$

$$rf_v = \frac{1}{ColNo} * \sum_{i=1}^{ColNo} P_{(v,i)} \quad 3.11$$

where *v* is the row number (*v*=1, 2, 3, … RowNo), ColNo is the number of columns of the image matrix.

- RF from colour images (RGB RF) are derived in the same way as greyscale ones taking into consideration the colour image channels. So, every attribute *rf* will be extracted by calculating the mean of every pixel value *P* from every image row for the three R, G, and B colour image layers:

$$RGB\ RF = [rf_{1RGB}, rf_{2RGB}, rf_{3RGB}, \ldots, rf_{RowNo\ RGB}] \quad 3.12$$

$$\mathbf{cf}_{v\,RGB} = \frac{\mathbf{1}}{\boldsymbol{ColNo}} * \sum_{\boldsymbol{i=1}}^{\boldsymbol{ColNo}} \left[\boldsymbol{P}_{v,\mathbf{i}(R)}, \boldsymbol{P}_{v,\mathbf{i}(G)}, \boldsymbol{P}_{v,\mathbf{i}(B)}\right] \quad 3.13$$

3.2.3 Column and Row Features (CRF)

This group is the combination of CF and RF together.

- Grey CRF:

$$\mathbf{Grey\ CRF} = [\mathbf{Grey\ CF}, \mathbf{Grey\ RF}] \quad 3.14$$

- RGB CRF:

$$\mathbf{RGB\ CRF} = [\mathbf{RGB\ CF}, \mathbf{RGB\ RF}] \quad 3.15$$

3.3 Feature Selection and Reduction Techniques

Exploiting these techniques allows the feature spaces to be reduced, which in turn will help classifiers to obtain the highest accuracy performance with fewer feature numbers (Hira and Gillies, 2015). Furthermore, the selection and reduction of features will avoid overfitting problems in classification (Bolon-Canedo, Sanchez-Marono and Alonso-Betanzos, 2015). There is a huge pool of algorithms available for selecting and reducing the size of feature space. After investigating several methods of feature selection and reduction, this research utilizes two algorithms for feature selection, ReliefF and mRMR, and The PCA method of feature reduction, because they are able to find the most informative features. Feature selection algorithms pick out some variables from the original features in the dataset. So that the feature selection algorithms select sub-features from the full number of those in the dataset (Wang, Wang and Chang, 2016). Meanwhile, feature reduction algorithms find a new dimension based on transforming two or more correlated features into new

uncorrelated features. The number of new features has to be fewer than the original number to avoid any classification overfitting (Jolliffe, 2002). This research also proposes two simple-process algorithms, RFFS and RSV, to reduce the number of features in the datasets.

3.3.1 **ReliefF**

This is one of the common and reliable feature selection methods. It is a filter-based method. ReliefF has been extended from Relief to solve the limitation of dealing only with binary classes (Durgaba, 2014). This algorithm works by finding the weights W

Begin

- set all weights W[A]:=0.0;
- for i:=1 to N (samples of dataset) do begin
  - randomly select an instance r_i;
  - find nearest hits $nh(r_i)$;
  - for each $Class \neq r_i$ do
  - from *Class* find nearest misses *nm*(Class);
  - for A:=1 *to* M (features of dataset) *do*
    - $W[A] = W[A] - \sum \frac{diff(A,r_i,nh)}{(nm \times nh)} + \sum_{Class \neq r_i} \frac{\left[\frac{p(Class)}{1-p(r_i)} \sum diff(A,r_i,nm)\right]}{(nm \times nh)}$

End

Figure 3-3 The ReliefF steps for calculating the weights of every feature of the dataset

of every feature according to the value of the nearest neighbour in the same class *r* and then the other classes. These steps are illustrated in Figure 3-3.

where; W is the weight of every attribute A, r_i is a random row from the dataset, N is the rows of the datasets (samples), *nh* is a selected instance from the same class, *nm* is a selected instance from different classes, M is the columns of the datasets (features)

3.3.2 Minimum Redundancy Maximum Relevance (mRMR)

(Peng, Long and Ding, 2005) proposed mRMR as a method for filtered feature selection. For a given x(S×F) where S represents classes and F features, mRMR aims to pick *NewFeatures*; where *NewFeatures* < F. The features picked strongly correlate with classes but they do not correlate among themselves. An intensive search is applied according to two common search methods using mutual information difference (MID) criterion and mutual information quotient (MIQ) to find the minimum redundancy and maximum relevancy:

$$\boldsymbol{MinumumW} = \frac{\mathbf{1}}{|\boldsymbol{s}|^{\mathbf{2}}} \sum_{\boldsymbol{i,j} \in \boldsymbol{s}} \boldsymbol{MI}(\boldsymbol{i,j}) \qquad 3.16$$

$$\boldsymbol{MaximumV} = \frac{\mathbf{1}}{|\boldsymbol{s}|} \sum_{\boldsymbol{i} \in \boldsymbol{s}} \boldsymbol{RI}(\boldsymbol{h,i}) \qquad 3.17$$

where $|s|$ is the number of classes in the dataset *S*, *MI(i,j)* is the mutual information between the i^{th} and j^{th} features of the class. *RI(h,i)* is the relevance between the feature *i* and the target class *h*.

3.3.3 Principle Component Analysis (PCA)

PCA applies an orthogonal linear transformation of the features to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on, in a given matrix *dataset*(N×M), where N is the number of samples (rows), and M is the number of features (columns) (Jolliffe, 2002). The mean of the training samples is given:

$$\boldsymbol{mean} = \frac{\mathbf{1}}{\mathbf{M}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{M}} \boldsymbol{x_i} \qquad 3.18$$

The matrix of *dataset* is centralized by subtracting *mean* from each x_i.

$$y_i = x_i - mean \quad (3.19)$$

Furthermore, the covariance matrix is estimated by:

$$cov = \frac{1}{M}\sum_{i=1}^{M} y_i \cdot y_i^T \quad (3.20)$$

Eventually, PCA calculates the eigenvectors and eigenvalues of the covariance matrix as well as selecting the top desired element number of eigenvectors that correspond to the top element numbers of eigenvalues.

3.3.4 Formulating Two Feature Selection Methods

The highly relevant features selected using ReliefF and mRMR in this research show that pixels with big values are more informative to help classifiers distinguish between classes in the dataset. Therefore, the proposed selection methods: the rectangular frame feature set (RFFS) and removing small values (RSV) will discard the pixels of the small values in two different ways.

3.3.4.1 Rectangle frame feature set (RFFS)

The proposed identification method allocates Eimeria oocysts in the centre of the images, which means that pixels in the exterior columns and rows of image matrices have small values because the colour of the background is black. RFFS will repeat the remove of the first and last columns and rows of image matrices, and these steps are explained in Figure 3-5. So, the dimensions of the image i(x,y) will be reduced in every round of RFFS which leads to decreasing the sizes of CF, RF, and CRF sets. Basically, this algorithm reduces the image dimension by cutting off the first and the last rows and columns every time. Then, the features of CF, RF, and CRF will be extracted from the new images of *i*. This process will be continued while the row and column numbers are larger than 3, as this algorithm cuts two rows and two columns each time. The number of the feature sets extracted from the images will be based on

Begin

- *i*(x,y) 'oocyst image after image processing stage'
- for all processed images
- While x>3 and y>3 do
  - New image=I(x-*2*, y-*2*); '*remove first and last rows and columns'*
  - From New image, Extract Grey: CF, RF, and CRF; RGB: CF, RF, and CRF

End

Figure 3-5 The algorithm of RFFS

the dimensions of the images. Every extracted feature set will be investigated in the

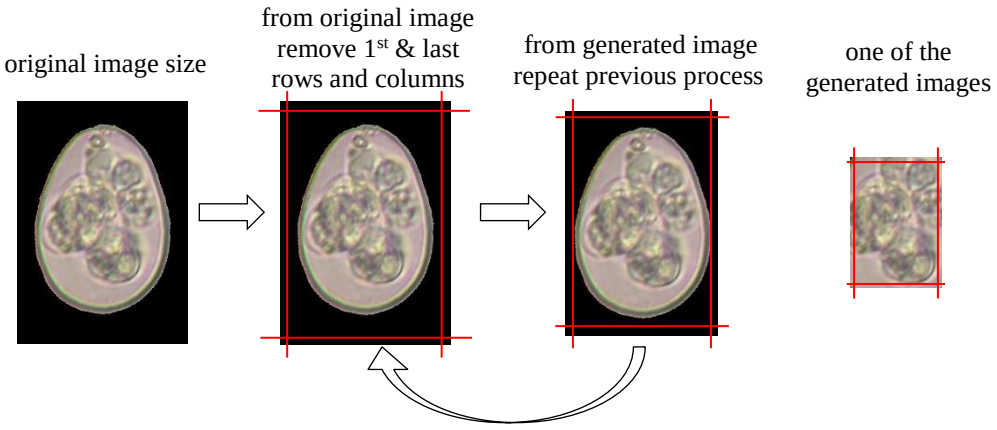


Figure 3-4 Framing oocyst images by removing two rows and columns to reduce the segmented features

classification stages. Figure 3-4 shows how the sizes of the images are decreased.

3.3.4.2 Removing small values (RSV)

RSV works on the vectors of CF, RF, and CRF extracted from grey and RGB datasets. As the small values of pixels are non-dominant distinguishing features for identification, therefore, this method aims to remove the small values from every sample in all classes of the dataset. Figure 3-6 shows the steps of RSV. First, it sorts features of every sample (row) in ascending order, and then it cuts out the features with small values according to the desired feature number. This method will be successful if it helps prediction models to obtain significant improvement in accuracy

performances with fewer features than the original number in every dataset. The threshold value is determined based on the required feature number (*#needed features)*. This algorithm will consider the features with the biggest values first, because it will sort them in descending order.

Begin

Let *data*(N×M) = one of the extracted datasets (Grey or RGB : CF, RF, or CRF); N is sample number; M is feature number

- for every sample (row=1, 2,3 ,… N*)* of *data* do
  - sort *data*(row,M) in descending order
- select *#needed features*
- let *selected_data= selected_data*(N×*#needed features*)

End

Figure 3-6 The algorithm of RSV

3.4 Classification

This is a process that assigns the objects considered to predetermined categories (Duda, Hart and Stork, 2000). This research exploits three of the supervised classification models to classify segmented pixel-based feature sets. K-NN, ANN, and SVM models are used in every experiment.

3.4.1 K-Nearest Neighbour (K-NN)

K-NN is a non-parametric classifier which does not make any assumptions about the underlying distribution of data. Its rule is to create K votes for the similar instances to unseen observations in the testing set. The value of K is not identical for all cases. Therefore, it can be empirically selected, but it is recommended that it should be an odd value so as to avoid confusing the classifier during the voting processes (Chomboon *et al.*, 2015; Plamenov, Nikolova and Georgieva, 2016). This research

applies K values from 1 to 19 to find which gives the results. K-NN has several distance metrics. The best results using K-NN is yielded when the cityblock distance, also called Manhattan is utilized (Kataria and Singh, 2013). The distance between two points is calculated by summation of the absolute difference of the Cartesian coordinates, as shown in the following equation.

$$\boldsymbol{d_{st}} = \sum_{\boldsymbol{i=1}}^{\mathbf{n}} \left| \boldsymbol{x_{sj}} - \boldsymbol{y_{tj}} \right| \qquad 3.21$$

3.4.2 Artificial Neural Networks (ANNs)

An ANN contains a sequence of layers. Every layer has a set of nodes/neurones. The neurones is every layer are linked by weighted connections to all neurones on the preceding and succeeding layers (Suzuki, 2011). Figure 3-7 depicts the general structure of the ANN. This research applies a feedforward two-layer NN with Bayesian regularization as a learning algorithm, because it helps the network to obtain the best classification results. The numbers of neurons in the input and output layers are equal to the number of dataset features *N* and observations *m* respectively. Every neuron in the input layer will receive one attribute for every observation in the dataset.

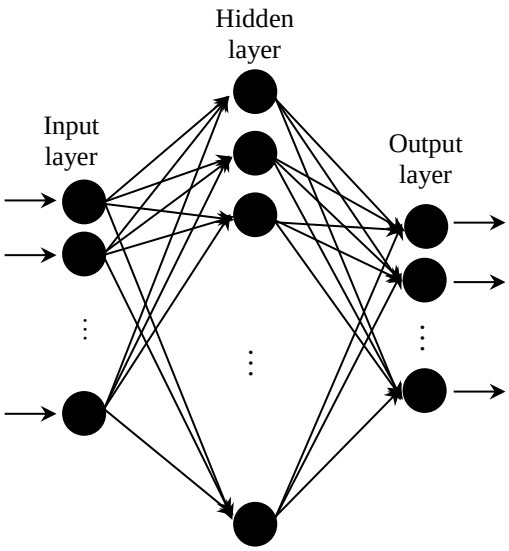


Figure 3-7 The general structure of ANN

Thereafter, those neurons will pass their values, after applying the adopted activation function to the hidden layer's neurons. There are several methods to determine the number of neurons in the hidden layer (Stathakis, 2009). This research applies the following equation to select the number of the neurons in the hidden layers:

$$\textbf{The neuron number of the hidden layer } = \textbf{INT}\left(\sqrt[2]{(\boldsymbol{M}+\mathbf{2})\boldsymbol{N}} \right) \qquad 3.22$$

Where INT is to convert to integer value; M is the number of the features (columns of the dataset); and N is the number of the samples (the rows of the dataset).

3.4.3 Support Vector Machine (SVM)

The support vector machine (SVM) was first built as a binary classifier to process linear and non-linear classification (Ertekin, 2009). More recently, it has been updated to be used as a multi-class classification model involved in various techniques to deal with multi-class problems such as one-versus-one, one-versus-all, or LibSVM tools (Shah, 2007; Sangeetha and Kalpana, 2011; Chang and Lin, 2013). This research utilizes SVM, because it is one of the common and effective classifiers (Ertekin, 2009). The main concept of SVM is to find perfect boundaries between observations. Two parameters are carefully selected to ensure that SVM yields its best result. They are regularization parameter (C) and the kernel width which called gamma (γ) (Amami, Ben Ayed and Ellouze, 2015). SVM has ability to create linear or non-linear boundaries. This discrimination is achieved by maximizing the separation margin and minimizing the error rate. This research utilizes LibSVM tool as multi-class SVM model with the Radial basis function (RBF) as a kernel classifier function.

3.4.4 Evaluation of Classification

Evaluating the performance of the classifiers is a very important issue in deciding whether or not the models applied are robust. The evaluation process starts with

finding an appropriate division of the dataset into training and testing sets (Duda, Hart and Stork, 2000), and then, precise calculations of the classification performance are required to ensure convincing results. There are several methods of evaluating the performance of prediction. This research applies the following techniques:

3.4.5 Cross-validation (CV)

CV is a statistical method which divides the dataset in different ways. This research uses a k-fold CV technique (Kohavi, 1995). This divides the dataset into k numbers of sub-datasets. The k-fold CV will create k-sub sets, where each contains almost the same number of observations or a very converged proportion. For example, if the dataset is divided into 5 folds; this means that every fold will contain almost 20% of the samples of the observations which form the dataset. In classification, k-1 folds will be used as a training set and one fold will be the testing set. Furthermore, it is necessary to give every fold a chance of being used as testing data in order to avoid any classification bias (Yang *et al.*, 2011). This research alternatively repeats the classification process by considering different folds as testing set. 5-fold CV is applied, and so every experiment is repeated 5 times.

3.4.6 Confusion Matrix

This is often a useful tool to display classification results in tabular form. Showing numbers of samples whether or not samples are correctly classified. Additionally, any misclassification results are clearly shown (Visa *et al.*, 2011).

3.4.7 Performance Measurements

According to the findings shown in the confusion matrices, classification performance is calculated in this research based on the average of the total number of correct classification/True positive (TP) from the whole population of the dataset samples. In

cases of unbalanced datasets which do not show equal numbers of samples in every class, the weighted average is recommended to precisely calculate the accuracy of classification performance; as shown in Eq 3.23:

$$\boldsymbol{Accuracy} = \frac{\mathbf{1}}{\boldsymbol{classes}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{\#classes}} \left(\frac{\boldsymbol{TP_i}}{\boldsymbol{samples}}\right) \quad 3.23$$

where *classes* is the number of classes, TP is the true positive (the number of the correct classification) of class *i*, and *samples* is the actual population of class *i*.

3.4.8 Standard Deviation (SD)

Standard deviation (SD) is a quantity that expresses the plus or minus differences between a group of values and their average. This research repeats every experiment with the random creation of 5-fold CV 50 times. Therefore, the SD will show how robust the yielded results are.

$$\boldsymbol{SD} = \sqrt{\frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{values}} (\boldsymbol{X_i} - \boldsymbol{\mu})^{\mathbf{2}}}{(\boldsymbol{values} - \mathbf{1})}} \quad 3.24$$

where values is the number of the results obtained (the accuracy rate obtained from every experiment), (X_i - μ) is the difference between every result and the average of all results.

3.5 Finding Super-Pixel Feature Set

Several studies have pointed out that super-pixels might be selected based on their colour, position, intensity, or geometric characteristics using various methods (Achanta *et al.*, 2011; Akbar *et al.*, 2015; Cong *et al.*, 2015). This research utilizes ReliefF feature selection to seek the super-pixel set based on colour. Instead of CF, RF, and CRF sets, the images themselves will be driven into ReliefF. The size of super-pixels will be changed until high accuracy gains are obtained.

3.6 **Parasite Image Datasets**

Two microscopic image databases of protozoan parasites of the genus Eimeria have been acquired from (Castañón B *et al.*, 2007). The microscopic images were captured under an optical microscope with 400X magnification by cameras of 4-megapixels and the resolution of those images is 2272×1704 pixels with 300 dots per inch (dpi). Then those images were saved as 24-bit JPEG (Castañón *et al.*, 2007). Those two databases were partially analysed by the same authors. Every image of these two databases were cropped out from the whole images of the microscopic slides. The publishers discarded some images because of the mis-segmentation of oocysts. In contrast, the image processing phase in this research has performed better, increasing the number of the correct detection of oocyst images from the published dataset. This research analyses these two databases which are provided online and well organized in separate folders. Therefore, the analysis of these images based on the proposed methodology in this research will consider two challenges. The first is to analyse more images. The second is to obtain better accuracy compared with the results of the publishers.

As mentioned in the CH2, the Eimeria species are microscopic objects and ovals, but not regular circles (their sizes are 9 to 34 µm in length and 6 to 21 µm in width). The publishers investigated three groups of the morphological features, but the contribution of the curvature group, which contained three features mean of curvature, curvature standard deviation, and entropy curvature, was less than 1%. This means that despite the complexities of extraction of the morphological features, the results might not be improved. Therefore, the hypothesis of utilizing pixel-based features to classify the Eimeria species from the microscopic images might provide high results based on uncomplicated features.

As discussed in Ch 2 also, the analysis of the microscopic images is the first way to detect and distinguish Eimeria species, but the extreme similarities between the Eimeria species often force veterinaries to rely on the other methods: dissection or DNA test. The proposed methodology in this research analyses the medical microscopic images of Eimeria species utilising pixel-based features rather than the morphological features to provide quick automatic identification.

3.6.1 Chicken Image Database

A database of 4484 microscopic images of the seven chicken Eimeria species has been published online (Castañón B *et al.*, 2007), whilst the authors only analysed 3891 images as shown in Table 3-1. The proposed image processing and segmentation stages have succeeded in segmenting more of images. Figure 3-8 depicts images the Eimeria species from the dataset analysed in this research.

Table 3-1 Numbers of images in chicken Eimeria species database

Eimeria species	Class Label	Number of images published in (Castañón B *et al.*, 2007)	Number of images used in (Castañón *et al.*, 2007)	Number of images used in this PhD study
E. Acervulina	ACE	744	636	736
E. Brunetti	BRU	442	418	437
E. Maxima	MAX	360	321	353
E. Mitis	MIT	825	757	813
E. Necatrix	NEC	519	404	493
E. Praecox	PRA	898	747	887
E. Tenella	TEN	696	608	683
Total Number		4484	3891	4402

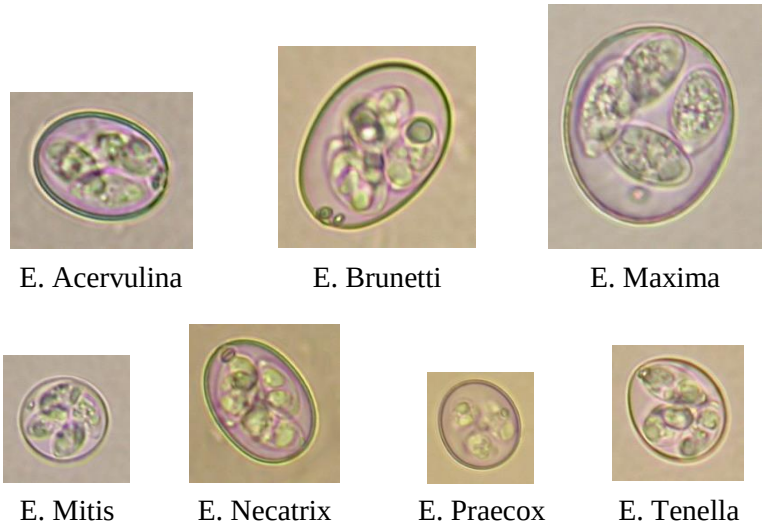


Figure 3-8 Samples from the image database of the Eimeria species in chickens

3.6.2 Rabbit Image Database

An image database of 3237 microscopic images of the eleven Eimeria species in rabbits has been published online (Castañón B *et al.*, 2007). The authors only utilized 2167 images out of the total number. However, the proposed study could accurately detect 2902 oocyst images. Table 3-2 below shows the numbers of images of every species in the rabbit images. Samples of the Eimeria species in rabbits are shown in Figure 3-9.

Table 3-2 Numbers of images in rabbit Eimeria species database

Eimeria species	Class Label	Number of images published in (Castañón B *et al.*, 2007)	Number of images used in (Castañón *et al.*, 2007)	Number of images used in this PhD study
	COE	283	191	229
E. Exigua	EXI	292	282	292
E. Flavescens	FLA	521	127	413
E. Intestinalis	INT	170	186	134
E. Irresidua	IRR	213	209	184
E. Magna	MAG	572	291	556
E. Media	MED	315	199	292
E. Perforans	PER	210	110	198
E. Piriformis	PIR	136	133	136
E. Stiedai	STI	229	156	180
E. Vejdovskyi	VEJ	296	283	288
Total Number		3237	2167	2902

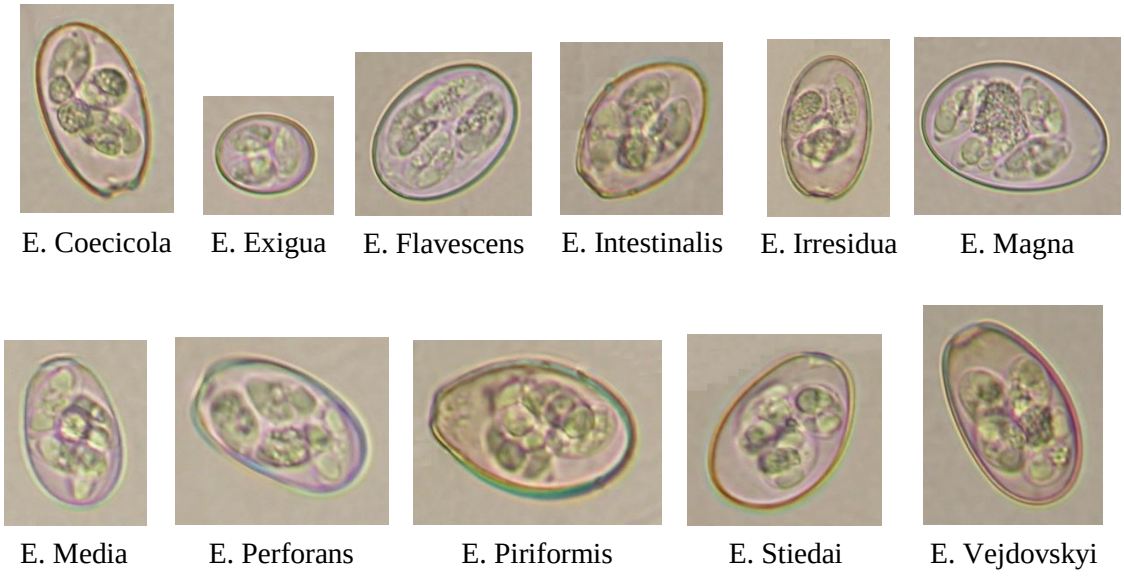


Figure 3-9 Samples from the image database of the Eimeria species in rabbits

3.6.3 The Analysis of the Discriminative Features Used to Identify the Oocysts of the Eimeria Species

The morphological characteristics have been widely analysed as main descriptors to distinguish between the oocysts of the Eimeria species form the microscopic images. Veterinaries in laboratories analyse the morphological features to classify the Eimeria species from the microscopic images. Furthermore, the automatic identification studies rely on the same morphological features to distinguish between these oocysts of the Eimeria species from the microscopic images.

Despite the extreme similarities between these oocysts, they can be barely recognized from the microscopic images utilizing the measurement of their sizes, shapes, symmetries, and textures. The lengths and widths of the oocysts of the Eimeria species are ones of the important discrimination features, but the lengths and widths of the oocysts are often overlapped between the species which creates classification errors. Table 3-3 and Table 3-4 show the range of the lengths and widths of the oocysts of the Eimeria species in chickens and rabbits, respectively. For example, the length of E.

Brunetti is from 20.7 to 30.3 µm. This range is overlapped with E. Maxima, E. Necatrix, and E. Tenella. Additionally, the width is also overlapped in several species, such as E. Mitis and E. Necatrix. Moreover, the sizes of oocysts of the rabbit Eimeria species show overlapping. For instance, the lengths of E. Intestinalis, E. Magna, and E. Piriformis are very similar and overlapped. Additionally, the widths of E. Exigua, E. Intestinalis, and E. Perforans are overlapped. Furthermore, there is a strong similarity in the length and width between E. Coecicola and E. Vejdovskyi. All these similarities and overlapped sizes create difficulties to distinguish between the oocysts of the Eimeria species.

In addition, the oocysts found in the feces might be unsporulated or sporulated. Both cases are infective. Therefore, the texture characteristics cannot often be informative (Zonaed Siddiki *et al.*, 2014). The internal bodies in the case of sporulated oocysts are called sporozoites. They might consist of 2, 4, 6, or 8 sporozoites for all species. So, counting these sporozoites will not help to provide informative features. Furthermore, clustering the sporozoites is not useful too, because they do not remain in a specific place inside the oocysts (Zonaed Siddiki et al., 2014). Other features such as the areas and perimeters of the oocysts are common features to distinguish Eimeria species. Although the colour of the oocysts of Eimeria species tends to the yellow colour, the analysis of colours from the pixel values based on the approach of Border/Interior Pixel Classification (BIC) showed important results (Saito *et al.*, 2015).

Table 3-3 The range of the length and the width of Eimeria species in chickens (Saif, 2008)

Eimeria Species	Length	Width
E. Acervulina	17.7 – 20.2 µm	13.7 – 16.3 µm
E. Brunetti	20.7 – 30.3 µm	18.1 – 24.2 µm
E. Maxima	21.5 – 41.5 µm	16.5 – 29.8 µm
E. Mitis	11.7 – 18.7 µm	11.0 – 18.0 µm
E. Necatrix	13.2 – 22.7 µm	11.3 – 18.3 µm
E. Praecox	19.8 – 24.7 µm	15.7 – 19.8 µm
E. Tenella	19.5 – 26.0 µm	16.5 – 22.8 µm

Table 3-4 The range of the length and the width of Eimeria species in rabbits (Oliveira et al., 2011)

Eimeria Species	Length	Width
E. Coecicola	24.5 – 29.3 µm	11.3 – 15.4 µm
E. Exigua	12.3 – 14.6 µm	12.8 – 15.8 µm
E. Flavescens	23.6 – 25.8 µm	14.7 – 17.0 µm
E. Intestinalis	19.4 – 25.2 µm	12.4 – 14.6 µm
E. Irresidua	26.3 – 28.4 µm	21.6 – 23.7 µm
E. Magna	21.8 – 24.2 µm	15.6 – 18.3 µm
E. Media	28.7 – 30.5 µm	16.0 – 19.2 µm
E. Perforans	18.3 – 22.4 µm	10.9 – 14.2 µm
E. Piriformis	22.7 – 26.4 µm	15.6 – 17.2 µm
E. Stiedai	27.8 – 32.4 µm	17.2 – 19.4 µm
E. Vejdovskyi	25.2 – 29.11 µm	14.3 – 16.0 µm

Eventually, the morphological characteristics including various types of features have been utilized in different studies. Despite satisfactory results obtained based on analysing morphological descriptors as discussed in Ch 2, the extraction of those descriptors requires very lengthy processes. In several cases some of those features might not be informative; such as the curvature features (Castañón *et al.*, 2007). Therefore, the extraction of very informative and uncomplicated features is needed to provide a reliable identification of Eimeria species. This research analyses the pixels of the oocysts to identify Eimeria species. The pixel values represent all the oocyst morphological features. So, instead of the extraction of the morphological characteristics, this research assumes that utilizing the pixel values directly from the oocysts of Eimeria species after unifying their orientation and positions will provide

simple extracted as well as very informative features. As discussed previously in this Chapter. The pixel-based features are extracted calculating the mean of every column and row of the pixels from oocyst image matrices. Considering one case to explain how pixel-based features can be very informative to identify the oocysts of the Eimeria species, the classification of the most important feature selected by ReliefF algorithm from the RGB CRF set using ANN showed that the analysis of one feature can clearly help the classifier to differentiate between the Eimeria. Figure 3-10 and Figure 3-11 depict the range of the values of the most important feature selected from RGB CRF using ReliefF technique for chickens and rabbits, respectively. The figures show that although there is clear overlapping between the values of the feature selected, the pixel-based features are useful to help the classifier to classify the Eimeria species specially when the overall accuracy rates significantly increased with analysing more features. The following chapters discuss the results yielded from all pixel-based features, the selection of features, and the super-pixels.

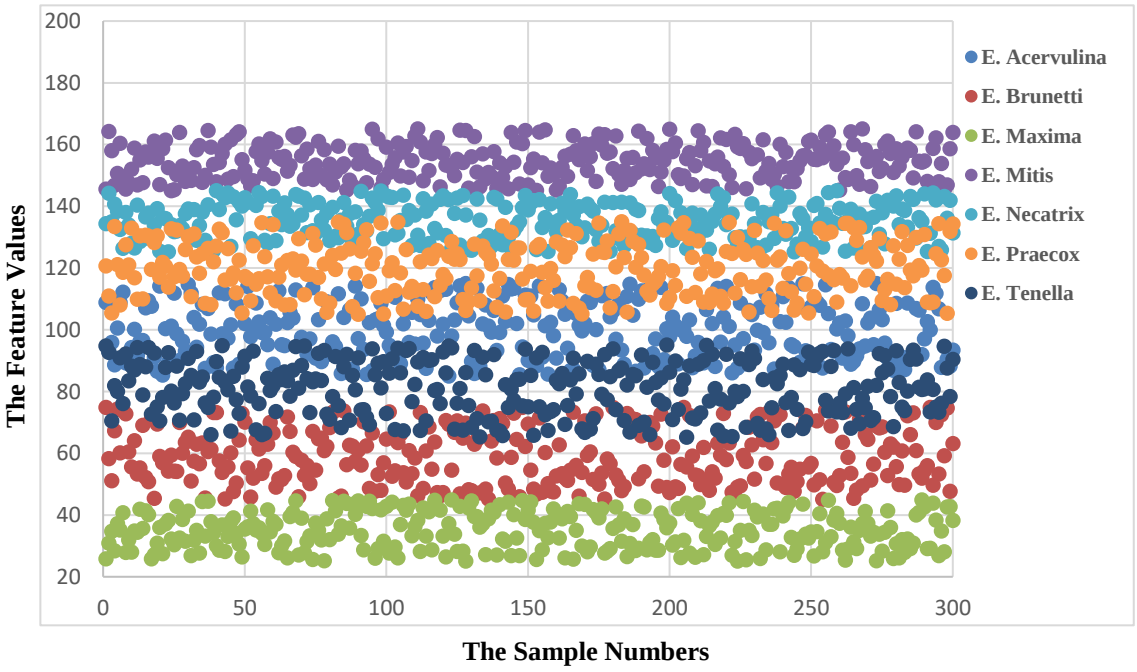


Figure 3-10 The range of the values of the most important feature selected from RGB CRF of 2100 samples (300 samples for every class) of Eimeria species in chickens using ReliefF algorithm.

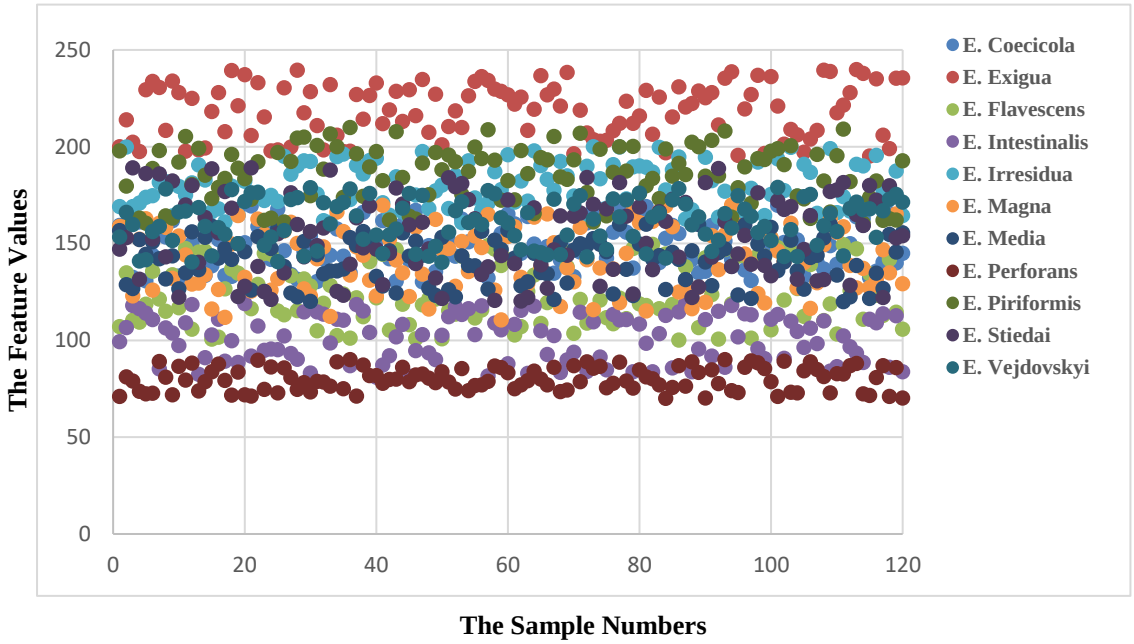


Figure 3-11 The range of the values of the most important feature selected from RGB CRF of 1320 samples (120 samples for every class) of Eimeria species in rabbits using ReliefF algorithm.

3.7 **Summary**

This chapter has described the stages of the methodology used in this research to analyse and identify microscopic images of protozoan parasites in the genus Eimeria. According to the proposed methodology, this research examines pixel-based feature sets in contrast to the numbers of previous studies, which relied on various groups of morphological characteristics. The present research hypothesis is that these simple extracted features from the microscopic images can be very reliable and informative features in distinguishing between Eimeria species. This chapter has also discussed the databases of chicken and rabbit images used. These databases are analysed using the proposed identification method in the following chapters.

CHAPTER 4 CLASSIFICATION RESULTS FOR COLUMN, ROW, AND COLUMN+ROW FEATURE SETS

This chapter includes the results obtained for the classification of parasite types for both chickens and rabbits by using entire feature set and the three extracted feature sets including the column feature (CF), row features (RF), and the combined column+fow feature (CRF). The same feature sets were extracted for both the original size and the reduced size of the images to assess if image re-sizing has any effect on classification performance. This is also further explored for both grey-level and colour images. As mentioned in Chapter 3, the classification was achieved by using K-NN, ANN, and SVM methods to assess the accuracy and stability of different classification algorithms. In order to avoid a bias towards the random selection of subsets of the images for 5-fold cross-validation, the same process was repeated 50 times, and average accuracy levels and corresponding standard deviations are given in the tables. The overall accuracy rates are calculated from the results shown in the confusion matrices. The results in the confusion matrices show in details the number of correct or incorrect classification results i.e. true positive (TP), true negative (TN), false positive (FP), or false negative (FN). Furthermore, the overall accuracy rate will be calculated by finding the weighted average of the TP results of every class from the confusion matrix as described in Chapter 3. The confusion matrices of all experiments are listed in Appendix A, B, and C.

4.1 Results for K-Nearest Neighbour Classifier

Odd K values of 1 to 19 have been assigned to find out the best possible results. Several distance measures have been applied to obtain the highest results, but the most significant outcomes for both image databases have been yielded by the city-block distance metric, which is also called Manhattan. The results yielded by K-NN varied and all showed a small level of variation during the 50 runs.

4.1.1 Results from the Chicken Image Database

The investigated chicken image database contains 4402 microscopic images. It has been analysed based on the proposed method divided into two groups: grey level (grey CF, RF, CRF) and colour (RGB CF, RF, and CRF) sets.

Table 4-1 illustrates the significant results obtained from analysing chicken image database by using K-NN classifier in both grey and colour datasets. The results from RGB set are better than the grey sets. Furthermore, CRF sets always gives better results than CF, and the RF sets always provide the lowest overall accuracy rates among greyscale and colour datasets.

Table 4-1 The highest overall accuracy from chicken datasets by K-NN

	Grey sets			RGB sets		
	CF	RF	CRF	CF	RF	CRF
μ	82.14%	79.45%	83.34	93.79%	89.07%	94.02%
σ	0.58%	0.53%	0.55%	0.42%	0.09%	0.53%
K	9	15	13	5	7	1

The highest results among all of the datasets have been calculated by finding the weighted average of correct classification rates based on the generated confusion matrices. The most significant correct classification results in term of using K-NN to classify the whole attribute numbers of the all of chicken sets have been obtained from

RGB CRF feature set. Table 4-1 shows that the results gained are quite reliable due to the small values of standard deviation of less than ±1% in every case. The CF dataset shows a clear improvement from grey to colour results which for approximately 12.5% higher. Similarly, the CRF dataset achieved 10.68% higher accuracy in the case of the colour dataset.

Although the result for the RGB RF set is 10% higher than grey RF dataset, comparing these results with those of the CF and CRF sets could lead to the conclusion that the RF dataset does not contribute much to improving CRF results when K-NN was applied.

4.1.2 Results from the Rabbit Image Database

Using the K-NN classifier to classify all of features of the feature set of rabbit dataset groups yielded diverse various results. Similar to the results from the chicken image database, the RGB datasets gave correct classification rates higher than for the grey ones. The RGB CRF dataset provided the best results. Table 4-2 shows the results obtained from classification of the rabbit datasets using K-NN. These results are the highest correct classification rates, which have been calculated from the generated confusion matrices.

Table 4-2 Highest overall accuracy for rabbit datasets using K-NN

	Grey sets			RGB sets		
	CF	RF	CRF	CF	RF	CRF
μ	70.82%	67.5%	80.73%	83.88%	79.61%	84.37%
σ	1.55%	1.09%	1.4%	1.38%	1.05%	1.29%
K	5	11	9	9	11	5

These results have shown improvements up to 14% in the overall accuracy rates for colour datasets compared to greyscale level. For instance, the classification of the grey CF dataset yielded 70.82% (±1.55%) as an overall accuracy rate, whilst the RGB CF set provided 83.73% (±1.38%). Although merging the grey RF into grey CRF

increased the classification accuracy by approximately 10%, its role was still very weak in the case of RGB CRF at less than 0.5%.

Despite a slight increase in values of standard deviation compared with the results from the chicken feature sets, the results from rabbit images yielded by K-NN are still reliable because the values of the standard deviation values remain smaller than ±1.5%.

4.2 Results for Artificial Neural Network

Neural networks consist of various hidden units based on the number of input variables with Bayesian regularization has been exploited as a classification model to identify Eimeria species in both image databases. In general, the outcomes from the use of ANN are significantly higher than those for K-NN.

4.2.1 Results from the Chicken Image Database

Classification of the seven Eimeria oocyst species using ANN provided results higher than those using K-NN. Table 4-3 depicts the overall accuracy values obtained from every dataset. These overall results have been calculated based on the correct classification rates of every true positive value in the confusion matrices. The best overall accuracy rate has been yielded from RGB CRF of 95.57% (±0.79). Appendix B shows all of the confusion matrices of the chicken datasets when ANN was applied. Additionally, the values of standard deviation indication that the dispersion of 50 iteration for all of grey and RGB datasets are still close to the overall average values which means that the classification results are reliable.

Table 4-3 The overall accuracy for chicken sets using ANN

	Grey datasets			RGB datasets		
	CF	RF	CRF	CF	RF	CRF
μ	83.53%	79.6%	85.41%	94.19%	90.68%	95.57%
σ	0.97%	1.07%	0.69%	0.82%	0.95%	0.79%

4.2.2 Results from the Rabbit Image Database

ANN achieved significant classification results with all of the grey and colour rabbit datasets when it was applied to classify the full number of the features in the datasets. These results are shown in

Table 4-4 the best result average was for RGB CRF at 94.21% (±1.79%). The improvement overall progression from grey to RGB sets is quite similar to K-NN progression. For greyscale level datasets the results are lower than those for colour and the highest result is for the CRF, then CF, but the results for RF sets are always the lowest in both the greyscale and RGB sets. The accuracy rates have been calculated based on the confusion matrices.

Appendix B contains the confusion matrices of every dataset of rabbits using ANN. The results gained by ANN from the rabbit feature sets show a slight increase in standard deviation values which, indicates that there are small variations among these results which are higher than those for the chicken images. However, the maximum value of standard deviation is still less than ±2.6%.

Table 4-4 The overall accuracy for rabbit sets using ANN

	Grey datasets			RGB datasets		
	CF	RF	CRF	CF	RF	CRF
μ	76.49%	71.98%	83.11%	91.34%	89.47%	94.21%
σ	2.39%	2.11%	2.24%	2.11%	2.57%	1.79%

4.3 Results for Support Vector Machine Classifier

Originally, the SVM classification model was built as a binary classifier, but it has recently been developed in different ways to be used as a multi-class classifier. LibSVM is one of the most commonly used tool for multi-class SVM prediction models. The SVM kernel function utilized in this research is Radial Basis Function (RBF). As described in Chapter 3, the RBF is an appropriate technique because it can

process the non-linear classification problems and it also has a strong ability to process a high-dimensional space, which this research deals with. Multi-class SVM with the LibSVM tool and RBF as a kernel function obtained significant results which are reported below. For both datasets, SVM obtained its best results when 1 and 0.0001 were assigned to the parameters of C and γ respectively. Only the best results yielded are shown in chapter; the other findings using SVM are shown in appendix C.

4.3.1 Results from the Chicken Image Database

Classification of the feature sets of chicken Eimeria species using the SVM has provided remarkable results. These results are depicted in Table 4-5. Similar to the other classifier models, the classification of RGB features gave accuracy rates higher than for grey features, with accuracy rates approximately 10% higher for colour features. The results gained by SVM reached 95.41% (±0.39%) for RGB CRF. This accuracy rate has been calculated based on the overall results from the true positive rate of classification outcomes.

Furthermore, the contribution of RF represents only small values in both groups, with increases of 3.72% and 0.53% for grey and RGB groups respectively. From the small values of the standard deviation, it could be stated that the classification of the chicken datasets using the SVM yielded results with less dispersion rates compared with the results for the other classifiers.

Table 4-5 Overall accuracy rates for chicken sets using SVM

	Grey datasets			RGB datasets		
	CF	RF	CRF	CF	RF	CRF
μ	82.87%	81.22%	86.50%	94.88%	91.62%	95.41%
σ	0.27%	0.29%	0.31%	0.24%	0.33%	0.39%

4.3.2 Results from the Rabbit Image Database

Table 4-6 shows the results for the rabbit Eimeria datasets using the SVM. The grey group yielded accuracy results lower than the RGB sets. The highest correct classification rate of 94.21% (±1.79%) has been obtained from RGB CRF. However, the grey RF provided a very low result 71.98% (±2.11%). The combination of grey RF with Grey CF to establish the grey CRF was able increase accuracy by 5.1%. Similarly, with RGB group, RF provided the lowest accuracy and also its role was very weak when merged with RGB CF set. Regarding the dispersion of the results, the maximum value of standard deviation have was ±2.57% for RGB results. This indicates that the results obtained using SVM do not show huge fluctuations among the 50 run times.

Table 4-6 Overall accuracy rates of rabbit sets using SVM

	Grey datasets			RGB datasets		
	CF	RF	CRF	CF	RF	CRF
μ	78.01%	73.47%	80.29%	92.06%	86.36%	93.03%
σ	0.95%	0.68%	0.9%	1.23%	1.08%	1.82%

4.4 Effect of Resizing Images on the Extracted Features

The image sizes have been reduced by exploiting three image resizing algorithms: Nearest, Bilinear, and Bicubic. Therefore, new image dimensions have been generated to represent: two-thirds, half, one-third, one-quarter, one-fifth, and one-tenth of the original size of the images. The proposed methods of the extraction of feature sets has been applied to every new image generated. Classification of the extracted features after applying these three minimizing algorithms by the considered classifier models have produced significant results. Apart from the one-fifth and one-tenth, types of

resizing algorithms do not badly affect the accuracy rates of gained results from the new created images. Moreover, the largest four sizes from every minimizing algorithms: two thirds, half, one third, and one-quarter have provided high results.

4.4.1 Results from Minimizing Chicken Image Size

The original size of the region of interest in the chicken image database is represented by image matrices of 510-by-370. The reduction of these images is shown in Table 4-7.

Table 4-7 New pixel sizes of chicken images after minimizing the original images

Resizing Degree	Image Pixels
Two-thirds	340-by-247
Half	255-by-185
One-third	170-by-123
One-quarter	128-by-93
One-fifth	102-by-74
One-tenth	51-by-37

Classification of the resized images up to one-quarter of the original image dimensions succeeded in presenting significant results. Use of different resizing image algorithms resulted in no obvious differences between them. As with the classification of results the segmented features from original images, the RGB CRF dataset always provided the highest accuracy, followed by the RGB CF. However, the RF of greyscale level or RGB images gave the lowest outcomes. The results yielded by K-NN, ANN, and SVM classifiers respectively when the images were minimized by Nearest interpolation. As can be seen in Table 4-8, 5-15, and 5-16, the classification results of the datasets minimalized in different degrees. For the one-quarter image size, RGB CF which contains 279 features and RGB CRF which contains 663 features gave accuracy rates of 92.56% (±1.06%) and 94.76% (±1.0%) when classified using ANN. Moreover, classification of this dataset using K-NN and SVM also provided highly accuracy results. However, the outcomes worsened sharply when the image dimensions were

reduced to less the one-quarter. Appendix D shows all of the classification results from minimizing the image database using all of the algorithms and the classifier models considered.

Table 4-8 Overall accuracy rates for classification chicken image sets when Nearest Neighbour interpolation was used to minimize image sizes

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (340×247)	K-NN	μ	83.58%	79.99%	85.07%	93.19%	90.32%	94.13%
		σ	2.02%	1.8%	4.89%	2.61%	1.96%	2.04%
	ANN	μ	88.07%	79.43%	88.98%	92.49%	88.15%	94.41%
		σ	1.23%	2.18%	1.45%	1.25%	1.75%	0.77%
	SVM	μ	81.65%	81.18%	85.05%	94.03%	90.21%	94.0%
		σ	2.26%	1.35%	1.88%	2.34%	1.7%	1.86%
½ (255×185)	K-NN	μ	83.72%	78.94%	84.57%	92.42%	89.53%	94.18%
		σ	1.62%	1.44%	5.38%	2.47%	2.01%	2.37%
	ANN	μ	86.96%	79.29%	89.34%	92.79%	88.22%	94.24%
		σ	1.17%	0.97%	1.32%	1.49%	1.74%	1.14%
	SVM	μ	81.8%	81.2%	84.41%	91.59%	90.91%	94.09%
		σ	1.6%	2.15%	2.15%	2.66%	1.83%	2.2%
⅓ (170×123)	K-NN	μ	83.03%	79.82%	84.67%	92.45%	88.68%	94.36%
		σ	2.3%	1.67%	4.76%	1.91%	2.04%	1.94%
	ANN	μ	86.15%	79.25%	88.75%	92.79%	87.8%	94.17%
		σ	1.38%	1.46%	1.32%	1.09%	1.77%	0.96%
	SVM	μ	81.47%	80.64%	84.54%	92.93%	90.37%	93.26%
		σ	2.21%	1.79%	2.35%	2.76%	1.65%	2.39%
¼ (128×93)	K-NN	μ	83.88%	79.89%	85.86%	92.12%	89.66%	94.13%
		σ	1.6%	1.58%	5.29%	2.13%	2.36%	2.0%
	ANN	μ	88.46%	80.55%	88.81%	92.56%	88.69%	94.76%
		σ	1.06%	1.64%	1.38%	1.06%	0.98%	1.0%
	SVM	μ	81.75%	81.28%	84.0%	92.94%	90.46%	93.62%
		σ	1.85%	1.86%	2.46%	2.76%	1.76%	2.16%
$^1/_5$ (102×74)	K-NN	μ	59.34%	58.0%	61.76%	63.5%	61.8%	64.28%
		σ	±3.02%	±3.64%	±3.5%	±3.09%	±3.84%	±3.2%
	ANN	μ	59.23%	57.8%	62.53%	65.48%	60.4%	67.3%
		σ	±2.20%	±3.21%	±3.08%	±3.69%	±3.8%	±3.3%
	SVM	μ	59.89%	57.3%	60.09%	60.8%	59.7%	61.74%
		σ	±2.70%	±3.84%	±3.5%	±3.46%	±3.70%	±3.65%
$^1/_{10}$ (51×37)	K-NN	μ	15.89%	13.2%	17.6%	18.78%	14.18%	19.4%
		σ	±5.23%	±6.2%	±4.8%	±4.78%	±5.9%	±4.84%
	ANN	μ	19.7%	18.71%	20.8%	22.8%	19.73%	24.9%
		σ	4.3%	4.68%	3.84%	3.61%	4.33%	3.78%
	SVM	μ	19.28%	17.36%	21.89%	21.92%	20.79%	22.9%
		σ	3.48%	3.54%	3.5%	3.47%	3.4%	3.61%

Table 4-9 Overall accuracy rates of classification for chicken image sets when the Bilinear algorithm was used to minimize the image sizes

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (340×247)	K-NN	μ	84.29%	79.3%	85.53%	94.55%	90.13%	93.47%
		σ	1.56%	1.56%	2.33%	2.33%	2.01%	2.16%
	ANN	μ	87.36%	79.03%	88.77%	92.33%	88.05%	93.98%
		σ	1.64%	1.93%	2.25%	1.47%	1.97%	0.78%
	SVM	μ	82.68%	81.37%	85.75%	93.28%	90.3%	94.08%
		σ	1.82%	2.03%	1.98%	2.39%	1.72%	2.48%
½ (255×185)	K-NN	μ	82.8%	79.97%	84.03%	94.08%	90.77%	93.17%
		σ	1.54%	2.03%	2.66%	2.27%	2.14%	2.07%
	ANN	μ	86.63%	79.25%	88.85%	92.75%	88.32%	94.07%
		σ	1.59%	2.08%	2.52%	1.62%	1.4%	0.81%
	SVM	μ	81.3%	81.21%	84.88%	93.24%	91.27%	94.14%
		σ	1.75%	1.75%	2.14%	2.47%	1.82%	2.46%
⅓ (170×123)	K-NN	μ	83.34%	79.99%	84.03%	93.04%	90.07%	93.32%
		σ	1.85%	1.86%	2.63%	2.07%	2.37%	2.29%
	ANN	μ	85.67%	79.02%	87.93%	92.49%	88.18%	93.92%
		σ	2.0%	1.89%	2.29%	1.71%	1.78%	1.03%
	SVM	μ	81.99%	81.6%	84.39%	92.56%	90.87%	92.88%
		σ	2.09%	1.93%	2.26%	2.89%	1.57%	1.9%
¼ (128×93)	K-NN	μ	82.53%	79.99%	83.82%	92.72%	90.09%	93.41%
		σ	1.79%	1.86%	2.38%	2.42%	2.52%	2.11%
	ANN	μ	88.08%	79.53%	88.53%	93.69%	88.04%	94.27%
		σ	1.27%	1.46%	2.27%	1.9%	2.02%	0.77%
	SVM	μ	82.71%	80.77%	84.06%	93.6%	90.48%	93.86%
		σ	1.77%	2.06%	1.82%	2.31%	2.04%	2.26%
$^1/_5$ (102×74)	K-NN	μ	63.2%	60.68%	64.08%	64.69%	62.79%	61.21%
		σ	3.49%	3.54%	3.47%	3.49%	3.52%	3.63%
	ANN	μ	58.93%	57.96%	62.53%	65.48%	60.4%	67.3%
		σ	2.92%	2.52%	2.66%	3.08%	3.39%	2.93%
	SVM	μ	59.27%	58.05%	60.51%	61.27%	59.24%	62.26%
		σ	2.77%	2.41%	2.9%	2.7%	2.87%	2.53%
$^1/_{10}$ (51×37)	K-NN	μ	20.45%	16.98%	21.4%	21.84%	19.68%	20.81%
		σ	4.7%	4.98%	4.83%	5.06%	5.01%	4.99%
	ANN	μ	20.11%	19.47%	20.98%	21.99%	19.61%	23.39%
		σ	4.7%	4.61%	4.63%	4.73%	4.55%	4.78%
	SVM	μ	19.7%	16.9%	21.6%	21.2%	20%	23.02%
		σ	3.94%	3.96%	4.09%	4.01%	4.08%	3.96%

Table 4-10 Overall accuracy rates of classification for chicken image sets when the Bicubic algorithm was used to minimize the image sizes

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (340×247)	K-NN	μ	85.13%	79.91%	85.15%	94.2%	91.12%	94.73%
		σ	2.58%	1.78%	2.37%	2.25%	2.17%	1.97%
	ANN	μ	87.31%	79.21%	88.81%	91.68%	88.17%	94.06%
		σ	1.91%	1.95%	2.25%	2.17%	2.59%	1.28%
	SVM	μ	81.84%	81.02%	85.%	94.6%	91.29%	94.16%
		σ	1.96%	1.84%	1.99%	2.48%	1.96%	2.27%
½ (255×185)	K-NN	μ	82.48%	79.25%	84.26%	94.11%	90.51%	94.66%
		σ	1.78%	1.37%	2.91%	2.02%	1.81%	2.12%
	ANN	μ	87.08%	79.39%	88.93%	92.4%	89.28%	94.%
		σ	1.59%	2.04%	2.39%	1.98%	1.96%	1.51%
	SVM	μ	81.41%	80.61%	84.72%	93.05%	91.06%	93.55%
		σ	2.04%	2.08%	2.12%	2.37%	1.97%	2.38%
⅓ (170×123)	K-NN	μ	83.97%	79.88%	84.63%	94.03%	90.57%	94.1%
		σ	2.19%	1.95%	2.9%	2.38%	2.33%	1.73%
	ANN	μ	87.36%	79.43%	87.46%	92.42%	87.18%	93.31%
		σ	1.66%	1.88%	2.65%	2.09%	2.39%	1.81%
	SVM	μ	81.85%	81.48%	84.73%	93.34%	90.52%	93.75%
		σ	1.73%	1.69%	2.21%	2.38%	2.05%	2.26%
¼ (128×93)	K-NN	μ	83.03%	79.32%	84.56%	92.92%	90.81%	93.77%
		σ	2.15%	1.44%	2.72%	2.16%	1.98%	2.45%
	ANN	μ	88.23%	79.71%	88.75%	93.51%	88.18%	93.27%
		σ	2.13%	2.03%	2.22%	2.03%	2.14%	1.57%
	SVM	μ	81.85%	79.97%	84.66%	92.89%	90.83%	93.25%
		σ	1.97%	1.86%	2.25%	2.7%	1.81%	2.53%
1/5 (102×74)	K-NN	μ	59.18%	59.4%	61.0%	62.84%	60.79%	63.96%
		σ	3.91%	3.91%	3.93%	3.94%	4.08%	3.88%
	ANN	μ	60.04%	58.1%	62.01%	65.8%	59.08%	65.87%
		σ	3.36%	4.29%	3.8%	4.0%	4.6%	3.64%
	SVM	μ	60.02%	57.87%	60.3%	60.18%	57.95%	61.0%
		σ	3.13 %	3. 68%	3.34%	3.66%	3.18%	3.55%
1/10 (51×37)	K-NN	μ	16.49%	15.3%	19.4%	19.3%	15.3%	19.77%
		σ	4.82%	5.1%	4.89%	5.09%	4.83%	4.88%
	ANN	μ	19.98%	18.6%	20.49%	21.97%	19.0%	24.3%
		σ	4.08%	4.45%	3.9%	4.04%	4.07%	3.96%
	SVM	μ	19.9%	18.3%	20.9%	21.66%	20.1%	22.76%
		σ	3.92%	3.98%	3.96%	3.94%	3.96%	3.95%

4.4.2 Results from Minimizing Rabbit Image Size

The region of interest of rabbit Eimeria species images is represented by dimensions of 482-by-335. Minimizing rabbit images by the same applied rates as with the chicken images has provided new image sizes as shown in Table 4-11.

Table 4-11 New pixel sizes of rabbit images by minimizing the original images

Resizing Degree	Image Pixels
Two-thirds	321-by-223
Half	241-by-168
One-third	161-by-112
One-quarter	121-by-84
One-fifth	96-by-67
One-tenth	48-by-34

The prediction results for the extracted feature sets from these new image sizes has shown that the accuracy rates obtained from the first four large image dimensions are significant results. These results are quite similar to those obtained based on the classification of the extracted datasets with the original image size. On the other hand, the attribute number has been reduced but nonetheless the accuracy rates are still high. However, classification of the datasets with images one-fifth and one-tenth the size of the original yielded very poor results. As with the chicken image database, there is no clear change in results when using any particular image resizing algorithm.

The best results for the one-quarter size set are 92.88% (±2.05%), 93.69% (±1.85%), and 93.52% (±2.12%) using the Nearest, Bilinear, and Bicubic image resizing algorithms respectively when ANN was applied.

Table 4-12, 4-13, and 4-14 list the highest accuracy results from different image sizes. Appendix D contains all the classification results from the minimized rabbit images.

Table 4-12 Overall classification accuracy rates for rabbit image sets when the Nearest Neighbour algorithm was used to minimize image sizes

Minimizing Degree (pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (321 × 223)	K-NN	μ	69.64%	63.39%	79.87%	83.11%	78.33%	82.84%
		σ	0.67%	1.55%	2.1%	2.16%	5.39%	2.25%
	ANN	μ	76.34%	71.33%	82.73%	91.19%	89.86%	93.05%
		σ	4.01%	2.31%	5.07%	3.36%	4.78%	1.75%
	SVM	μ	77.62%	71.08%	82.9%	92.02%	85.6%	92.49%
		σ	1.02%	1.23%	2.29%	1.99%	2.6%	2.15%
½ (241 × 168)	K-NN	μ	69.87%	63.36%	79.47%	82.54%	77.13%	82.92%
		σ	0.61%	1.23%	2.34%	2.46%	5.66%	2.49%
	ANN	μ	76.61%	71.41%	80.56%	91.26%	88.64%	93.98%
		σ	3.09%	3.21%	5.01%	3.13%	4.01%	2.46%
	SVM	μ	77.65%	71.07%	82.66%	91.84%	85.11%	92.82%
		σ	0.93%	1.54%	2.07%	1.82%	2.69%	1.85%
⅓ (161 × 112)	K-NN	μ	69.91%	63.49%	79.28%	82.96%	77.89%	82.62%
		σ	0.63%	1.63%	2.35%	2.75%	5.24%	2.23%
	ANN	μ	74.67%	70.73%	80.54%	90.36%	88.31%	93.42%
		σ	3.47%	3.25%	5.02%	2.81%	4.82%	2.84%
	SVM	μ	77.7%	70.99%	82.65%	92.25%	85.77%	92.65%
		σ	0.97%	1.61%	2.26%	2.32%	3.07%	1.72%
¼ (121 × 84)	K-NN	μ	69.94%	62.71%	79.34%	82.74%	77.72%	82.71%
		σ	0.64%	1.42%	2.09%	2.6%	5.67%	2.82%
	ANN	μ	74.8%	70.06%	80.23%	90.0%	88.95%	92.88%
		σ	3.78%	2.11%	4.66%	3.29%	5.48%	2.05%
	SVM	μ	77.63%	70.74%	82.43%	91.99%	84.65%	92.82%
		σ	0.83%	1.72%	2.44%	2.11%	2.72%	1.86%
1/5 (96 × 67)	K-NN	μ	45.98%	43.20%	47.50%	49.30%	47.62%	52.46%
		σ	3.56%	4.4%	3.51%	3.24%	4.25%	4.47%
	ANN	μ	47.6%	45.3%	49.8%	51.73%	49.78%	53.5%
		σ	4.52%	4.20%	4.60%	4.9%	4.82%	4.93%
	SVM	μ	47.1%	45.8%	50.1%	51.74%	48.87%	54.19%
		σ	2.77%	2.83%	2.54%	2.89%	2.9%	3.12%
1/10 (48 × 34)	K-NN	μ	19.80%	16.20%	22.85%	22.40%	21.97%	25.60%
		σ	6.26%	7.1%	6.12%	6.25%	6.87%	6.2%
	ANN	μ	22.5%	20.8%	23.58%	23.8%	21.2%	27.7%
		σ	5.85%	6.0%	5.6%	6.25%	6.82%	7.25%
	SVM	μ	20.48%	18.25%	22.85%	23.5%	19.4%	26.4%
		σ	3.25%	3.89%	3.65%	3.20%	3.85%	3.56%

Table 4-13 Overall classification accuracy rates from rabbit image sets when the Bilinear algorithm was used to minimize image sizes

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (321 × 223)	K-NN	μ	70.25%	64.84%	79.57%	83.32%	76.39%	83.81%
		σ	2.73%	2.08%	1.8%	2.89%	6.08%	3.13%
	ANN	μ	77.06%	71.13%	83.06%	92.63%	89.52%	94.77%
		σ	2.82%	2.78%	5.44%	2.71%	4.3%	1.93%
	SVM	μ	77.7%	71.26%	82.89%	92.84%	85.47%	92.66%
		σ	0.86%	1.25%	2.48%	1.79%	2.87%	1.69%
½ (241 × 168)	K-NN	μ	69.82%	64.03%	79.36%	83.32%	77.45%	83.17%
		σ	2.63%	2.02%	2.18%	2.89%	5.68%	3.17%
	ANN	μ	76.14%	71.76%	81.3%	91.99%	89.41%	93.96%
		σ	3.36%	2.89%	5.1%	3.27%	5.03%	2.05%
	SVM	μ	77.89%	70.18%	82.4%	92.33%	85.91%	92.29%
		σ	1.01%	1.55%	2.59%	2.01%	2.72%	1.59%
⅓ (161 × 112)	K-NN	μ	69.32%	63.77%	78.98%	83.32%	75.93%	82.31%
		σ	2.55%	1.91%	1.94%	2.89%	7.2%	2.66%
	ANN	μ	75.61%	71.49%	80.63%	91.29%	87.37%	93.45%
		σ	3.04%	2.51%	4.61%	3.84%	4.6%	1.97%
	SVM	μ	77.92%	70.87%	81.89%	92.16%	84.56%	92.97%
		σ	0.78%	1.05%	2.32%	1.75%	3.04%	2.02%
¼ (121 × 84)	K-NN	μ	69.05%	62.94%	79.73%	81.84%	74.54%	81.61%
		σ	2.88%	2.1%	2.08%	2.09%	4.77%	3.18%
	ANN	μ	75.44%	70.12%	80.84%	90.23%	87.32%	93.69%
		σ	2.88%	3.25%	5.13%	3.03%	4.04%	1.85%
	SVM	μ	77.96%	70.85%	81.63%	91.83%	85.59%	92.43%
		σ	0.95%	1.42%	2.01%	1.74%	2.67%	1.88%
1/5 (96 × 67)	K-NN	μ	45.98%	44.3%	47.76%	49.3%	46.28%	51.67%
		σ	3.11%	3.8%	3.4%	3.18%	3.87%	4.11%
	ANN	μ	47.51%	45.25%	49.44%	51.24%	49.08%	53.50%
		σ	4.20%	3.89%	4.0%	3.76%	4.1%	4.83%
	SVM	μ	47.5%	45.2%	49.4%	52%	48.2%	55.6%
		σ	2.18%	3.2%	2.43%	2.22%	3.1%	3.4%
1/10 (48 × 34)	K-NN	μ	20.02%	17.08%	21.45%	21.95%	21.05%	25.56%
		σ	5.94%	6.73%	6.23%	5.8%	6.23%	5.92%
	ANN	μ	21.93%	19.94%	23.59%	23.71%	22.08%	26.86%
		σ	5.70%	5.98%	5.11%	5.87%	6.3%	6.92%
	SVM	μ	21±13%	18.79%	23.24%	23.35%	20.83%	26.37%
		σ	3.55%	3.62%	3.18%	3.56%	3.47%	3.22%

Table 4-14 Overall classification accuracy rates from rabbit image sets when the Bicubic algorithm was used to minimize image sizes

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (321 × 223)	K-NN	μ	70.55%	64.84%	79.85%	82.84%	76.26%	83.98%
		σ	2.22%	2.08%	1.64%	2.01%	7.1%	3.41%
	ANN	μ	76.34%	71.32%	82.69%	91.75%	88.76%	93.96%
		σ	3.52%	2.61%	3.64%	3.15%	5.31%	2.36%
	SVM	μ	77.59%	71.46%	82.43%	92.36%	85.71%	92.83%
		σ	1.05%	1.5%	2.88%	1.97%	2.85%	1.81%
½ (241 × 168)	K-NN	μ	69.53%	64.03%	79.55%	82.42%	76.52%	82.14%
		σ	2.27%	2.02%	2.08%	1.64%	6.7%	2.05%
	ANN	μ	74.77%	70.94%	82.69%	90.92%	88.43%	93.58%
		σ	3.39%	2.88%	3.76%	3.9%	5.3%	2.33%
	SVM	μ	77.9%	70.78%	82.65%	92.43%	85.2%	92.91%
		σ	1.07%	1.42%	2.8%	1.37%	2.79%	2.05%
⅓ (161 × 112)	K-NN	μ	69.74%	63.77%	79.51%	82.73%	75.49%	83.32%
		σ	2.37%	1.91%	2.21%	1.92%	5.95%	2.89%
	ANN	μ	75.06%	70.63%	81.64%	89.82%	87.93%	93.72%
		σ	3.78%	3.3%	3.31%	3.44%	4.81%	1.62%
	SVM	μ	77.99%	70.31%	81.89%	91.91%	85.5%	92.89%
		σ	0.88%	1.62%	2.32%	1.97%	2.51%	1.84%
¼ (121 × 84)	K-NN	μ	69.45%	62.94%	79.79%	82.66%	76.35%	82.38%
		σ	2.51%	2.1%	2.0%	1.99%	5.75%	2.96%
	ANN	μ	74.81%	70.75%	81.57%	90.57%	88.36%	93.52%
		σ	3.09%	3.07%	3.89%	3.89%	5.33%	2.12%
	SVM	μ	77.91%	70.7%	82.19%	91.44%	84.78%	92.45%
		σ	0.83%	1.37%	2.29%	2.0%	2.66%	1.89%
1/5 (96 × 67)	K-NN	μ	45.98%	43.75%	47.63%	49.3%	46.95%	52.07%
		σ	3.37%	4.13%	3.76%	3.19%	4.17%	4.29%
	ANN	μ	46.98%	45.62%	48.98%	51.80%	49.76%	54.10%
		σ	3.99%	4.02%	4.04%	4.15%	4.54%	4.22%
	SVM	μ	47.19%	45.54%	49.90%	51.84%	48.94%	54.62%
		σ	3.57%	3.65%	3.28%	3.04%	3.74%	3.09%
1/10 (48 × 34)	K-NN	μ	19.91%	16.64%	22.15%	22.18%	21.51%	25.58%
		σ	4.29%	5.06%	4.44%	4.04%	4.49%	4.78%
	ANN	μ	21.0%	18.20%	23.30%	22.84%	21.01%	25.85%
		σ	5.60%	6.02%	5.48%	5.62%	6.25%	5.70%
	SVM	μ	20.87%	18.41%	23.13%	23.23%	20.41%	26.21%
		σ	3.67%	4.08%	3.17%	3.55%	4.00%	3.97%

4.5 Classification Results of Image Channel Features Individually

Classification of the features in the red, green, and blue image channels has been applied to find out which one of these channels is more important. Applying the proposed identification method on every image channel separately instead of normalizing them in a greyscale level or combining them to represent all channels together has resulted in differences in accuracy levels for the image channels in both of the databases.

4.5.1 Results from Analysis of Image Channels in Chicken Feature Sets

Table 4-15 illustrates the results yielded from the chicken image sets. The overall classification accuracy rates of these results were in between grey and RGB results. In other words, the overall results were better than those obtained from greyscale features, but worse than from RGB. In general, there are no great difference in the results obtained. However, the best overall accuracy rate was 89.65% (±0.25%) from the Columns of the Green Channel feature set using SVM classifier. In addition, Blue Channel features always provide the lowest accuracy rates.

Table 4-15 Overall classification accuracy for every image channel of the chicken database

Classifier		Columns			Rows		
		Red	Green	Blue	Red	Green	Blue
K-NN	μ	88.21%	88.86%	80.29%	83.19%	83.31%	79.87%
	σ	0.44%	0.53%	0.89%	0.76%	0.78%	0.78%
ANN	μ	88.61%	88.99%	79.14%	81.53%	81.48%	76.32%
	σ	0.93%	0.78%	0.87%	0.92%	1.05%	1.42%
SVM	μ	89.22%	**89.65%**	81.72%	85.92%	86.61%	81.62%
	σ	0.25%	0.02%	0.39%	0.32%	0.37%	0.38%

4.5.2 Results from Analysis of Image Channels in Rabbit Datasets

Table 4-16 depicts the classification accuracy results. The results obtained from all channels were close and sometimes less than the results for grey sets. The highest overall correct classification rate is 78.93% (±0.71%) yielded by the SVM classifier from the Columns of the Blue Channel dataset.

Table 4-16 Overall classification accuracy for every image channel of rabbit image database

Classifier		Columns			Rows		
		Red	Green	Blue	Red	Green	Blue
K-NN	μ	76.6%	76.93%	77.47%	72.36%	73.68%	73.11%
	σ	1.55%	1.47%	1.44%	1.6%	1.69%	1.67%
ANN	μ	78.18%	78.57%	76.85%	74.71%	74.71%	73.73%
	σ	1.12%	0.89%	0.86%	0.97%	1.12%	1.09%
SVM	μ	78.76%	77.85%	**78.93%**	75.64%	72.95%	74.1%
	σ	0.91%	0.78%	0.71%	0.84%	0.83%	0.78%

4.6 The Implementation of Published Identification Methods

According to the related work discussed in CH 2, various types of the morphological characteristics were investigated to obtain the highest possible accuracy rate in every published study. This section discusses three of the identification methods to find out the level of the accuracy rate can be obtained when they analyse the datasets utilized in this research of the chicken and rabbit Eimeria species.

The first study by (Castañón *et al.*, 2007) conducted on analysis to extract features from the images. As previously explained, these features varied in terms of characteristics of the nature of the category: curvature, geometry, and symmetry characteristics. In consequence, 13 features were used based on these categories. An automatic recognition approach was built based on classifying the features using a

Bayesian classifier. The datasets were divided into training and testing sets in rates 10:90 to 90:10 as shown in Tables 4-17 and 4-18. The selection of samples from every dataset was randomly applied. The experiment was repeated 9 times depending on the division method used. Then the overall accuracy rates were calculated from all experimental results. Tables 4-19 and 4-20 show the overall confusion matrices of the classification results from both datasets. The confusion matrices show in details the number of correct or incorrect classification results i.e. true positive (TP), true negative (TN), false positive (FP), or false negative (FN). Furthermore, the overall accuracy rate will be calculated by finding the weighted average of the TP results of every class from the confusion matrix. The TP results are those shown in green colour background. Overall prediction rates were found of 80.24% (±7.94%) and 76.21% (±8.81%) for the chicken and rabbit databases respectively.

Table 4-17 Classification results of every experiment changing the training and testing datasets from the chicken image database

Proportion	Training set	Testing set	Accuracy rate
10:90	440	3962	39.36%
20:80	880	3522	55.78%
30:70	1321	3081	73.93%
40:60	1761	2641	79.51%
50:50	2201	2201	87.70%
60:40	2641	1761	92.12%
70:30	3081	1321	96.66%
80:20	3522	880	98.39%
90:10	3962	440	98.63%

Table 4-18 Classification results of every experiment changing the training and testing datasets from the rabbit image database

Proportion	Training set	Testing set	Accuracy rate
10:90	290	2612	33.69%
20:80	580	2322	51.86%
30:70	871	2031	66.93%
40:60	1161	1741	78.8%
50:50	1451	1451	83.4%
60:40	1741	1161	89.8%
70:30	2031	871	92.7%
80:20	2322	580	93.8%
90:10	2612	290	94.6%

Table 4-19 Confusion matrix of the overall findings of the study of 13 features from the chicken dataset

		ACE	**BRU**	**MAX**	**MIT**	**NEC**	**PRA**	**TEN**
ACE	μ	**87.01%**	0%	0%	4.4%	3.9%	0.66%	4.03%
	σ	**12.1%**	0%	0%	2.82%	1.92%	0.68%	3.55%
BRU	μ	0%	**87.39%**	9.7%	0%	0%	2.91%	0%
	σ	0%	**11.99%**	8.76%	0%	0%	1.43%	0%
MAX	μ	0%	2.89%	**97.11%**	0%	0%	0%	0%
	σ	0%	2.37%	**1.91%**	0%	0%	0%	0%
MIT	μ	2.63%	0%	0%	**88.54%**	5.76%	0.25%	2.82%
	σ	1.45%	0%	0%	**7.8%**	6.53%	0.45%	2.09%
NEC	μ	18.99%	0%	0%	11.%	**57.49%**	0.08%	12.44%
	σ	15.12%	0%	0%	2.44%	**26.7%**	0.27%	8.64%
PRA	μ	0.74%	9.37%	0%	0.64%	0%	**70.51%**	18.74%
	σ	0.36%	7.55%	0%	0.58%	0%	**12.7%**	10.68%
TEN	μ	2.18%	0%	0%	10.01%	13.73%	0.48%	**73.58%**
	σ	1.44%	0%	0%	6.66%	9.09%	0.36%	**11.96%**

Table 4-20 Confusion matrix of the overall findings of the study of 13 features from the rabbit dataset

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**63.88%**	0%	4.08%	1.47%	1.85%	0.79%	6.51%	0%	0.56%	3.28%	17.57%
	σ	**10.9%**	0%	2.66%	1.15%	0.92%	0.66%	3.87%	0%	0.52%	0.99%	3.62%
EXI	μ	0%	**97.68%**	0%	0%	0%	0%	0%	2.32%	0%	0%	0%
	σ	0%	**4.22%**	0%	0%	0%	0%	0%	0.85%	0%	0%	0%
FLA	μ	2.26%	0%	**82.77%**	1.43%	1.18%	3.3%	1.63%	0%	0.89%	3.44%	3.09%
	σ	0.49%	0%	**4.82%**	0.63%	0.29%	0.86%	0.79%	0%	0.28%	1.65%	1.47%
INT	μ	3.1%	0%	0%	**73.11%**	0%	0%	4.91%	2.7%	11.1%	0%	5.08%
	σ	1.22%	0%	0%	**7.4%**	0%	0%	1.54%	0.56%	3.61%	0%	2.82%
IRR	μ	2.32%	0%	1.2%	0%	**67.84%**	20.61%	0%	0%	0%	8.03%	0%
	σ	1.33%	0%	0.89%	0%	**7.62%**	5.56%	0%	0%	0%	3.77%	0%
MAG	μ	3.02%	0%	8.02%	0%	6.06%	**80.62%**	0%	0%	0%	2.28%	0%
	σ	1.55%	0%	2.21%	0%	1.64%	**2.06%**	0%	0%	0%	0.87%	0%
MED	μ	6.47%	0%	3.08%	1.45%	0%	0%	**69.55%**	4.92%	1.12%	4.74%	8.67%
	σ	2.%	0%	1.08%	0.53%	0%	0%	**4.98%**	1.24%	0.08%	1.58%	2.13%
PER	μ	0%	2.42%	0%	0%	0%	0%	5.53%	**92.05%**	0%	0%	0%
	σ	0%	0.99%	0%	0%	0%	0%	1.99%	**11.42%**	0%	0%	0%
PIR	μ	3.35%	0%	1.61%	9.85%	0%	0%	4.96%	0%	**74.88%**	0%	5.35%
	σ	1.25%	0%	1.08%	2.36%	0%	0%	1.9%	0%	**22.97%**	0%	1.66%
STI	μ	14.79%	0%	8.23%	2.18%	1.57%	1.94%	4.72%	0%	0%	**57.14%**	9.43%
	σ	4.91%	0%	3.76%	1.34%	0.82%	0.81%	1.42%	0%	0%	**13.66%**	2.39%
VEJ	μ	9.88%	0%	1.34%	1.35%	0%	0%	5.48%	0%	1.52%	1.67%	**78.74%**
	σ	2.28%	0%	0.33%	0.53%	0%	0%	1.52%	0%	0.69%	0.58%	**6.9%**

The second considered study applied active shape model (ASM) technique as a tool for feature extraction from images and has been investigated for the classification of fish parasites (Ali, Hussain and Man, 2015). It applied 45 landmarks then were reduced using PCA. The analysis that reported best performance was obtained when the size of feature space was reduced to 22 features. Additionally, they classified features by LDA, K-NN, MLP, and SVM, and MLP yielded the highest performance. Therefore, the implementation of this method on the datasets considered in this research started with the same landmark number then it has been reduced using the same feature reduction method (PCA). The 10-fold CV was used to create the training and testing sets. So, the process of training and testing is repeated 10 times. The final accuracy rates are the overall of all processes. For the classification, MLP has been utilized. Generally, the overall accuracy from features reduced by PCA during the 10 run times were very low. The best results were with the chicken dataset 73.21% from 32 features variables and 62.4% from 38 features with the rabbit dataset. The accuracy rates have been calculated from the confusion matrices in Tables 4-21 and 4-22.

Table 4-21 Confusion matrix of the overall findings the study of utilizing ASM features from chicken dataset

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**94.54%**	0%	0%	1.39%	3.51%	0.53%	0.03%
	σ	**9.69%**	0%	0%	2.15%	1.28%	1.15%	2.05%
BRU	μ	0%	**84.12%**	0.8%	0%	0.01%	9.93%	5.14%
	σ	0%	**2.63%**	0.16%	0%	0.05%	0.5%	0.47%
MAX	μ	0%	0.73%	**99.27%**	0%	0%	0%	0%
	σ	0%	1.16%	**0.59%**	0%	0%	0%	0%
MIT	μ	1.51%	0%	0%	**91.72%**	1.91%	4.45%	0.41%
	σ	0.8%	0%	0%	**3.45%**	2.32%	1.44%	0.94%
NEC	μ	6.88%	0.7%	0%	8.15%	**45.48%**	17.25%	21.53%
	σ	1.51%	2.15%	0%	0.54%	**6.22%**	0.9%	2.78%
PRA	μ	3.82%	0.39%	0%	4.53%	25.28%	**54.01%**	11.97%
	σ	0.09%	0.3%	0%	0.32%	2.47%	**12.72%**	5.47%
TEN	μ	4.97%	0.51%	0%	5.89%	32.83%	12.45%	**43.35%**
	σ	2.16%	3.31%	0%	0.14%	2.87%	2.62%	**6.9%**

Table 4-22 Confusion matrix of the overall findings of the study utilizing ASM features from rabbit dataset

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**33.72%**	3.75%	11.38%	2.36%	0.87%	0.36%	13.41%	0.39%	0.14%	4.96%	28.68%
	σ	**4.59%**	0.36%	1.58%	0.27%	1.%	0.77%	2.96%	0.32%	0.21%	0.89%	2.28%
EXI	μ	0%	**93.%**	0%	4.%	0%	0%	0%	3.%	0%	0%	0%
	σ	0%	**1.68%**	0%	0.6%	0%	0%	0%	0.55%	0%	0%	0%
FLA	μ	2.65%	0%	**67.96%**	0%	1.32%	12.18%	1.33%	0%	4.48%	1.99%	8.1%
	σ	1.44%	0%	**6.81%**	0.17%	0.72%	0.91%	0.46%	0%	0.24%	0.32%	0.66%
INT	μ	0%	0%	1.81%	**68.45%**	0%	0%	2.66%	5.13%	2.18%	0%	19.78%
	σ	0%	0%	0.36%	**7.49%**	0%	0%	0.73%	0.79%	0.89%	0%	1.21%
IRR	μ	4.6%	0%	10.63%	0%	**67.18%**	17.6%	0%	0%	0%	0%	0%
	σ	0.16%	0%	0.84%	0%	**4.82%**	0.8%	0%	0%	0%	0%	0%
MAG	μ	0.19%	0%	16.87%	0%	2.49%	**66.22%**	0%	0%	0.1%	14.14%	0%
	σ	0.04%	0%	0.58%	0%	0.33%	**3.92%**	0%	0%	0.09%	0.08%	0%
MED	μ	11.12%	1.37%	1.97%	1.5%	0%	0%	**58.49%**	3.53%	0.35%	0%	21.68%
	σ	0.44%	0%	0.2%	0.4%	0%	0%	**5.95%**	0.2%	0.8%	0%	1.91%
PER	μ	0%	3.45%	0%	7.11%	0%	0%	12.57%	**76.86%**	0%	0%	0.02%
	σ	0%	1.4%	0%	2.2%	0%	0%	3.49%	**8.62%**	0%	0%	0.7%
PIR	μ	0%	0.74%	11.04%	12.28%	0%	0%	3.4%	0%	**45.99%**	0%	26.56%
	σ	0%	0%	0.87%	0.99%	0%	0%	0.47%	0%	**5.8%**	0%	1.66%
STI	μ	12.12%	0%	18.29%	0.73%	3.9%	8.34%	0.43%	0%	0.57%	**42.43%**	13.2%
	σ	4.25%	0%	4.97%	0.56%	0.12%	1.5%	0.32%	0%	0.08%	**9.09%**	2.14%
VEJ	μ	13.19%	0%	5.15%	4.78%	0%	0%	7.72%	0%	2.32%	0.78%	**66.07%**
	σ	2.87%	0%	0.72%	1.36%	0%	0%	1.73%	0%	0.59%	0.39%	**12.18%**

The last one analyses 10 morphological features (Kalafi *et al.*, 2016). The features are Euler number, perimeter, area, area density, perimeter density, centre of bounding box, length of bounding box, width of bounding box and orientation of bounding box. In the classification phase, K-NN was applied and the best results were obtained when K=10 nearest neighbours. The classification performance was assessed by dividing the datasets using leave-one-out (LOOCV) as well as 10-fold CV techniques. As described in CH 2, the results from LOOCV were higher than those of 10-fold CV. Applying LOOCV was likely possible in that study, because the image number was only 80 images. In this research the training and testing processes would take a very long time due to the sizes of the datasets. The analysis of both datasets using the extraction of 10 features considering K nearest neighbour from 1 to 20 was conducted and the two different dataset division techniques provided the highest results compared with the previous two methods. However, K-NN when K=7 with LOOCV yielded results of approximately 1.5% higher than those using 10-fold CV. The correct prediction rates were 89.16% (±2.51%) and 81.47 (±8.9%) from the chicken and rabbit datasets respectively. These accuracy rates have been calculated from the confusion matrices shown in Tables 4-23 and 4-24.

Table 4-23 Confusion matrix of the overall findings of the study of utilizing 10 features from chicken dataset evaluated by LOOCV

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**96.79%**	0%	0%	0.33%	0.96%	0.36%	1.55%
	σ	**2.18%**	0%	0%	0.24%	0.39%	0.02%	0.65%
BRU	μ	0%	**93.01%**	4.33%	0%	0%	1.27%	1.39%
	σ	0%	**2.24%**	0.87%	0%	0%	0.24%	0.44%
MAX	μ	0%	0.04%	**99.96%**	0%	0%	0%	0%
	σ	0%	0.16%	**0.49%**	0%	0%	0%	0%
MIT	μ	0.97%	0%	0%	**94.06%**	2.69%	1.03%	1.24%
	σ	0.21%	0%	0%	**1.74%**	0.73%	0.45%	0.47%
NEC	μ	8.9%	0%	0%	6.06%	**76.27%**	3.75%	5.02%
	σ	1.4%	0%	0%	2.85%	**3.97%**	1.39%	1.23%
PRA	μ	2.17%	2.45%	0%	2.1%	4.49%	**83.96%**	4.83%
	σ	0.49%	0.61%	0%	0.54%	0.9%	**3.12%**	1.05%
TEN	μ	3.91%	0.04%	0%	4.38%	4.1%	7.49%	**80.07%**
	σ	0.38%	0%	0%	0.8%	1.03%	1.44%	**3.81%**

Finally, the identification of Eimeria species using these three methods provided various accuracy rates and they also took different amounts of times to extract informative features. Table 4-25 shows the average length of the feature segmentation in every method. The ASM approach took the longest time, where 11.3 (±0.88) seconds were needed to analyse each image. That means that the ASM method will approximately take 19 minutes to extract features from 100 images. Meanwhile the method of extracting 10 features, which provided the best results, needed the shortest time compared to other methods. It roughly required 15 minutes to extract features from 100 images. Moreover, the implementation of these three identification methods, and the proposed method, were done using the same PC specifications (CPU i7 Intel 7^{th} generation, RAM 32GB, OS windows 7, and the Matlab platform 16b).

Table 4-24 Confusion matrix of the overall findings of the study of utilizing 10 features from rabbit dataset evaluated by LOOCV

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**78.21%**	0%	2.97%	1.34%	1.81%	0.95%	5.%	0%	0.64%	3.55%	5.52%
	σ	**10.89%**	0%	1.26%	0.74%	0.73%	0.75%	1.71%	0%	0.31%	1.29%	1.51%
EXI	μ	0%	**98.43%**	0%	0%	0%	0%	0%	1.57%	0%	0%	0%
	σ	0%	**1.1%**	0%	0%	0%	0%	0%	0.9%	0%	0%	0%
FLA	μ	1.59%	0%	**84.68%**	1.14%	1.11%	3.1%	1.25%	0.74%	0.78%	3.%	2.62%
	σ	0.58%	0%	**9.56%**	0.58%	0.57%	1.18%	0.52%	0.5%	0.13%	1.06%	0.87%
INT	μ	2.94%	0%	0%	**80.78%**	0%	0%	3.53%	2.16%	6.75%	0%	3.85%
	σ	2.09%	0%	0%	**11.01%**	0%	0%	2.%	0.66%	1.66%	0%	2.%
IRR	μ	1.47%	0%	1.14%	0%	**75.31%**	15.97%	0%	0%	0%	6.11%	0%
	σ	0.95%	0%	0.55%	0%	**14.53%**	4.11%	0%	0%	0%	2.78%	0%
MAG	μ	2.46%	0%	5.66%	0%	4.87%	**85.35%**	0%	0%	0%	1.67%	0%
	σ	0.65%	0%	1.48%	0%	1.15%	**4.12%**	0%	0%	0%	0.43%	0%
MED	μ	3.84%	0%	2.63%	1.46%	0%	0%	**79.61%**	2.78%	0.79%	2.64%	6.25%
	σ	1.48%	0%	0.73%	0.32%	0%	0%	**9.44%**	0.4%	0.24%	1.16%	2.23%
PER	μ	0%	2.06%	0%	0%	0%	0%	3.41%	**94.53%**	0%	0%	0%
	σ	0%	0.88%	0%	0%	0%	0%	1.86%	**2.9%**	0%	0%	0%
PIR	μ	2.69%	0%	0.83%	7.25%	0%	0%	3.46%	0%	**81.93%**	0%	3.85%
	σ	2.84%	0%	1.01%	1.32%	0%	0%	0.1%	0%	**3.33%**	0%	0.74%
STI	μ	10.22%	0%	7.38%	3.17%	2.68%	3.04%	4.57%	0%	0%	**63.68%**	5.26%
	σ	2.78%	0%	2.15%	1.63%	1.32%	1.62%	1.01%	0%	0%	**20.26%**	2.14%
VEJ	μ	8.59%	0%	2.16%	3.%	0%	0%	5.14%	0%	3.19%	4.29%	**73.64%**
	σ	2.85%	0%	1.43%	0.5%	0%	0%	1.77%	0%	1.87%	2.35%	**10.82%**

Table 4-25 Average duration of feature extraction stage for every image using the three different methods

	The methods		
	13 Features	ASM method	10 features
Extraction time (in seconds/image)	10.16 (±0.40)	11.3 (±0.88)	9.1 (±0.25)

The extraction of the pixel-based features from both databases requires very short time as shown in Table 4-26 . In the best case, it took 3.6 (±0.42) seconds when analysing of the grey datasets for images of the one-quarter image sizes. In contrast, the proposed method takes 8 (±0.73) seconds as the longest time to extract the RGB features for one original size image. This run time is obviously shorter than that for morphological feature methods.

Table 4-26 Average duration of feature segmentation stage for every image using the proposed methods analysing the original size and the one-quarter images

	Original image size		One-quarter images	
	Greyscale	RGB	Greyscale	RGB
Extraction time (in seconds/image)	4.8 (±0.6)	8 (±0.73)	3.6 (±0.42)	5.3 (±0.52)

4.7 Conclusions of the Results

Examining the pixel-based feature sets to identify species of Eimeria has been discussed in this chapter. For sets of greyscale levels CRF provided the highest results. Similarly, for RGB sets, CRF provided the highest classification results compared with RGB CF and RGB RF. In addition, CRF for both greyscale and RGB provided the best results when they classified by ANN. Although results using SVM are lower than ANN, the differences between the accuracy rates using ANN and SVM are very small. In contrast,

the accuracy rates using K-NN are always the lowest for greyscale and RGB sets in both the chicken and rabbit images. CRF for the greyscale and RGB have also provided the highest classification results using ANN during the experiment of minimizing image sizes. The standard deviation values of all results obtained from K-NN, ANN, and SVM are very small.

Considering those images which were incorrectly classified in both datasets, there are two cases for the mis-classification. The first case is due to mistakes in the segmentation of images which simply leads to errors in classification. The database of Eimeria in chickens included 23 images which could not be correctly segmented. In the database of Eimeria in rabbits, there were segmentation errors with 43 images. Figure 4-1 shows samples of some mis-segmented images.

Figure 4-1 Samples of images with incorrect segmentation

The second case is due to classifier errors. The frequencies of classification errors were not identical with all classifiers in all experiments. There were noticeable differences in the accuracy rates of the different classifiers. However, some images were always assigned to the wrong class by all classifiers. Figure 4-2 shows samples of Eimeria images which always led to errors in classification tasks, even though they were correctly segmented. In total, the number of oocyst images which always had incorrect

classification were 12 and 38 images in the chicken and rabbit databases respectively. These images require more investigation.

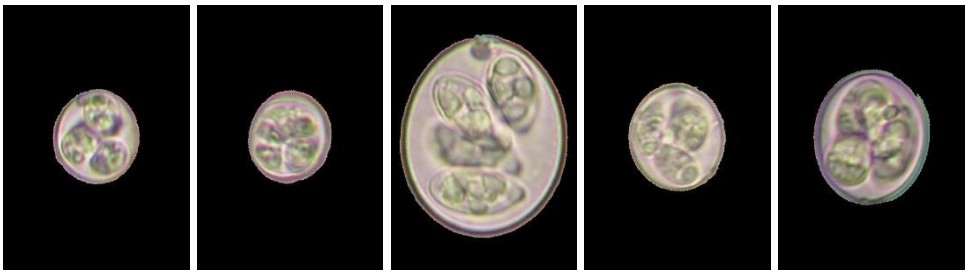


Figure 4-2 Samples of images incorrect classified

4.8 Summary

This chapter has shown the classification accuracy results of the extracted features based on analysis of the pixel values of Eimeria images from microscopic images, indicating that these features can be reliable in identifying Eimeria species. As discussed, various extraction ways have been to investigate pixel-based features. Furthermore, these features have been segmented form original image sizes and minimizing them up to one-tenth of the original size. The results described in this chapter have shown that pixel value features are reliable but give varying accuracy rates. The combinations of extracted features based on image columns and rows have provided the most accurate classification rates. In contrast, the results from RF sets always are less accurate in both grey and RGB sets. RGB features have always provided the better results. Moreover, reducing image sizes up to one-quarter did not significantly affect the accuracy of classification.

The results obtained from the three morphological methods discussed in this chapter showed that they are less than the results yielded from the pixel-based features. They are less in term of the overall accuracy rates and the time of the feature extraction.

Moreover, analysis of individually image channels did not reveal any particularly a reliable channel, because in the case of chicken image database the green channel features provided the best results followed by the red channel, and the blue channel results were always the lowest. However, for the rabbit image datasets the accuracy rates for different channels were not in that orders. The highest result gained was by with the

red channel, followed by the green, and then the blue channel. Therefore, the proposed method relies on an analysis normalizing these image channels at the greyscale level or combing them to represent RGB feature sets. In consequence, the classification of grey and RGB features can be adopted as a strong method to identify Eimeria species. Although this chapter has discussed the effect of minimizing image dimensions, which leads to a reduced feature space, features should be optimized by finding fewer features so as to avoid any overfitting classification problems. Hence, experiments testing whether or not these features can be optimized using different feature selection algorithms are discussed in the next chapter. Only RGB CRF feature set will be considered in the following chapter, because it provides the best overall accuracy rates with all classification models utilized compared with analysing CF and RF separately.

CHAPTER 5 APPLYING FEATURE SELECTION AND REDUCTION METHODS TO OPTIMIZE THE FEATURE SETS

The use of feature selection and reduction algorithms aims to help classifiers to avoid overfitting, this leads to minimizing confusion in classification as well as increasing the accuracy rates. This chapter presents and discusses the findings from the utilization of four feature selection methods namely ReliefF, mRMR, RSV, and RFFS and PCA as a feature reduction method to optimize the extracted features. The discussion in this chapter is only based on the results from the RGB CRF set in both image database and the performance results are shown in summary tables.

5.1 Application of Selection and Reduction Methods on RGB CRF of Chicken Datasets

As stated before, the microscopic images considered of chicken Eimeria species after the processing phase are represented in matrix image sizes of 510-by-370 pixels. Applying feature selection and reduction techniques will allow decreasing feature numbers as well as improving the classification accuracy.

The classification of the full feature space of RGB CRF yielded the highest overall accuracy rate among all the analysed datasets from all chicken image database. The results with RGB CRF presented in the previous chapter were good using all of the

prediction models considered. The optimization of this dataset using the adopted selection methods to detect the most relevant features leads to even better outcomes. Table 5-1 lists the best results obtained. ANN with ReliefF selection method yielded of an overall correct classification rate 96.70% (±0.89%) when 21% of the full size of feature space was processed. SVM yielded accuracy rate 0.8% lower, but the feature space used was only 15% of the original. K-NN with reliefF algorithm also achieved excellent results, providing 94.75% (±0.4%) accuracy from 742 variables, which represents approximately 28% of the original feature space. Furthermore, the results when features ranked by mRMR are lower than those for ReliefF, but they are still good and were quickly improved with a smaller feature space. For instance, K-NN results from mRMR-ranked features yielded 94.75% (±0.56%) accuracy when 1522 features were processed; however, the outcomes still reached 93% with 820 features. Figures 5-1, 5-2, and 5-3 depict the results achieved when widening the feature space.

Additionally, reducing the number features using the PCA method shows rapidly improving results, although these results often start with very low accuracy rates. In the case of K-NN classifier, the results with PCA reached an overall accuracy rate of 90% with 624 features. The improvements in the results with RSV and RFFS methods are still slow and need large size of feature dimensions.

Table 5-1 Highest overall accuracy rates of chicken RGB CRF by applying different feature selection methods and three different classifiers

Selection and Reduction Method		Classification Model		
		K-NN	ANN	SVM
ReliefF	Feature No	742	560	420
	Accuracy	94.75%	96.70%	95.9%
	SD	±0.4%	±0.89%	±0.47%
PCA	Feature No	2120	2343	1783
	Accuracy	93.99%	94.95%	94.81%
	SD	±0.43%	±0.78%	±0.6%
mRMR	Feature No	1522	960	940
	Accuracy	94.75%	95.6%	94.99%
	SD	±0.56%	±0.66%	±0.45%
RSV	Feature No	2300	1860	1862
	Accuracy	93.96%	95.1%	94.91%
	SD	±0.59%	±0.76%	±0.72%
RFFS	Feature No	1903	1204	1820
	Accuracy	94.77%	95.54%	94.74%
	SD	±0.44%	±0.81%	±0.58%

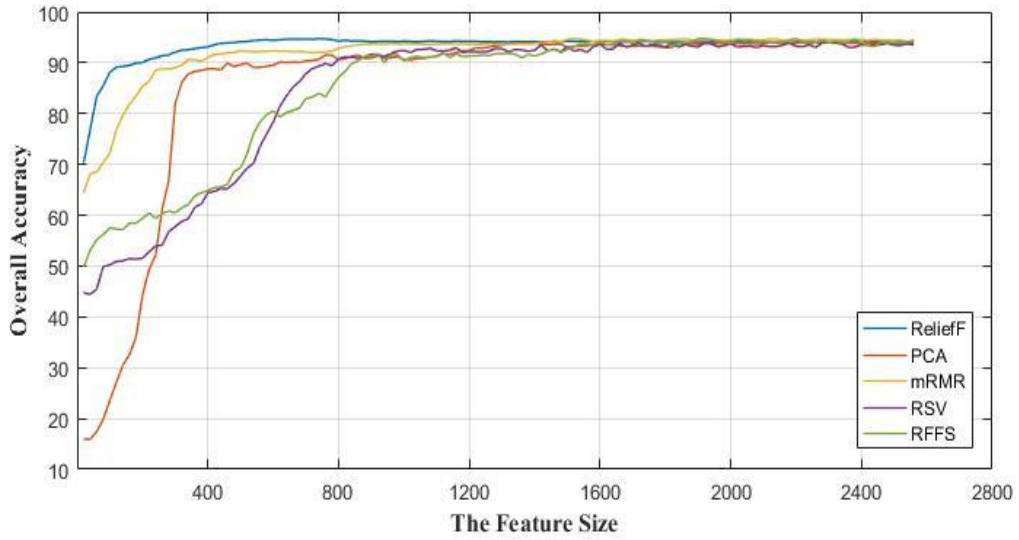


Figure 5-1 Improvement in overall accuracy rates by K-NN classifier while increasing the numbers of the selected features and applying different feature selection algorithms on chicken RGB CRF

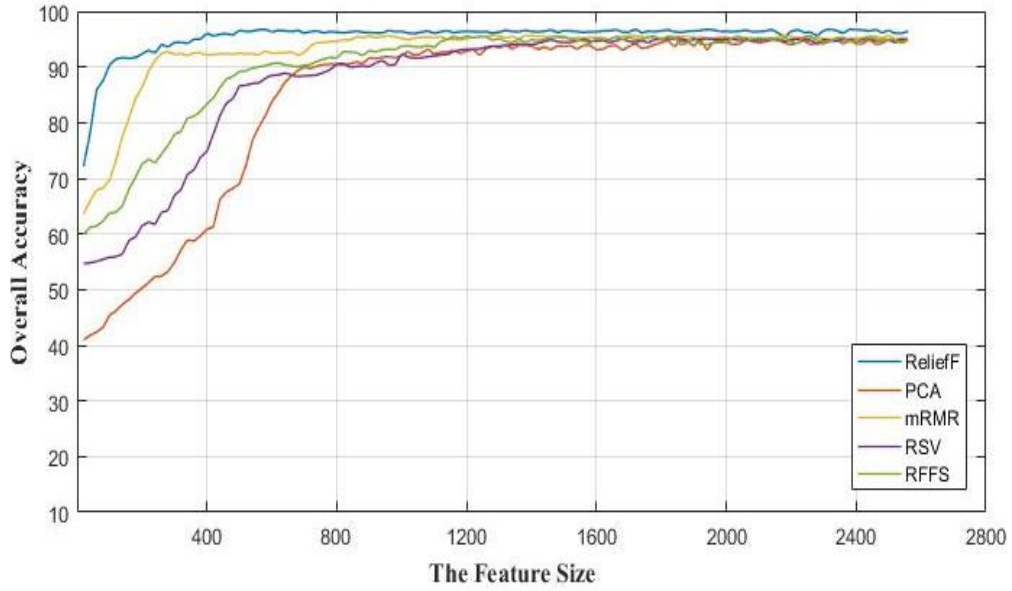


Figure 5-2 Improvement in overall accuracy rates by ANN classifier while increasing the numbers of the selected features and applying different feature selection algorithms on chicken RGB CRF

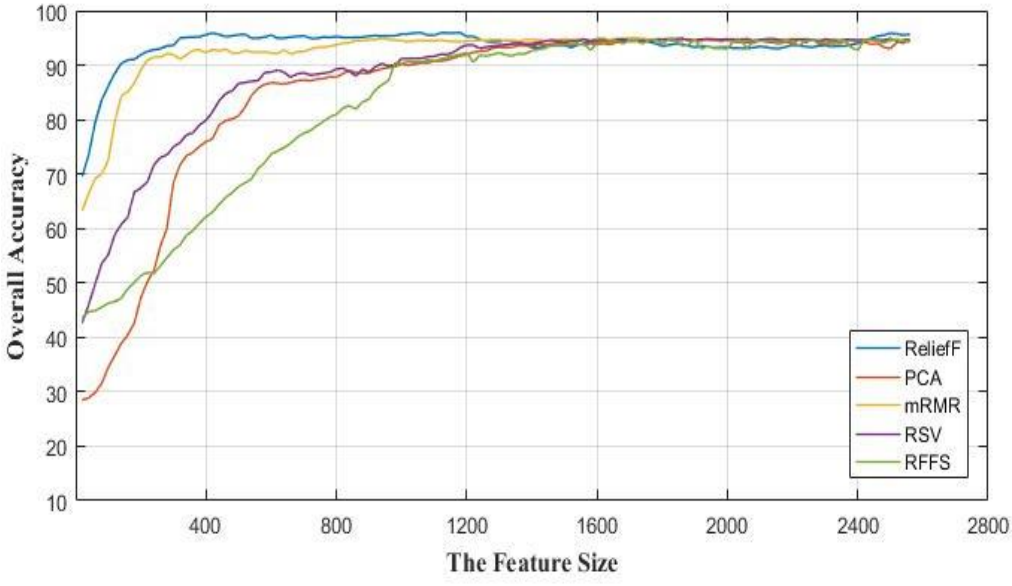


Figure 5-3 Improvement in overall accuracy rates by SVM classifier while increasing the numbers of the selected features and applying different feature selection algorithms on chicken RGB CRF

5.2 Application of Selection and Reduction Methods on RGB CRF of Rabbits Datasets

Classification results from the datasets extracted from this image database can be enhanced by applying the selection and reduction techniques. Originally, the processing stage has prepared the microscopic images of the rabbit database in image matrix dimensions of 482-by-335 pixels. Henceforth, the best performance results for rabbit datasets will be considered in terms of the smallest possible size of the feature space.

RGB CRF set contains 2451 features. In Chapter 4, the highest accuracy rate from classifying them all was 94.21% (±1.79%) using ANN. Table 5-2 shows the improvement in accuracy rate based on selecting the most relevant features of RGB CRF using different selection methods. Both ANN and SVM achieve the high performance with fewer features based on ReliefF ranking. The best overall accuracy rate of 95.85% (±2.4%) is yielded from analysis of the 992 most relevant features selected by ReliefF method and using ANN classification model. This feature space represents 40% of the total features. Furthermore, SVM yielded 94.73% (±1.26%) with fewer features using ReliefF selection method as well. ANN with RFFS selection method obtained 95.12% (±2.69%), but the feature space contained 1912 features, which constitutes 78% of the full dataset size.

Figures 5-4, 5-5, and 5-6 show the progress of the overall improvement in accuracy while adding features that are more relevant using the three selection methods. For the results using K-NN, ReliefF, mRMR, RFFS, and PCA methods all provided relevant features to increase the rates sharply with about 500 features, although K-NN outcomes are not the best. Furthermore, the results from K-NN with RSV showed very slow development even with more features. Additionally, ANN results include the highest accuracy results for RGB CRF dataset. The improvement curves are characterized by fluctuations, which means that there is some a variation in the results obtained. In spite of the ANN results showing noticeable dispersion, the standard deviation values do not exceed 3% in the worst cases.

Table 5-2 Highest overall accuracy rates of rabbit RGB CRF by applying different feature selection methods and three different classifiers

Selection and Reduction Method		Classification Model		
		K-NN	ANN	SVM
ReliefF	Feature No	1590	992	812
	Accuracy	85.67%	95.85%	94.73%
	SD	±1.59%	±2.4%	±1.26%
PCA	Feature No	2213	1912	1532
	Accuracy	84.09%	94.77%	93.74%
	SD	±1.02%	±2.72%	±1.99%
mRMR	Feature No	2232	1852	2252
	Accuracy	84.97%	94.96%	94.43%
	SD	±1.33%	±2.07%	±1.72%
RSV	Feature No	1554	1632	1112
	Accuracy	84.61%	94.94%	93.47%
	SD	±1.53%	±2.41%	±1.84%
RFFS	Feature No	1791	1912	1532
	Accuracy	85.00%	95.12%	93.48%
	SD	±1.28%	±2.69%	±1.86%

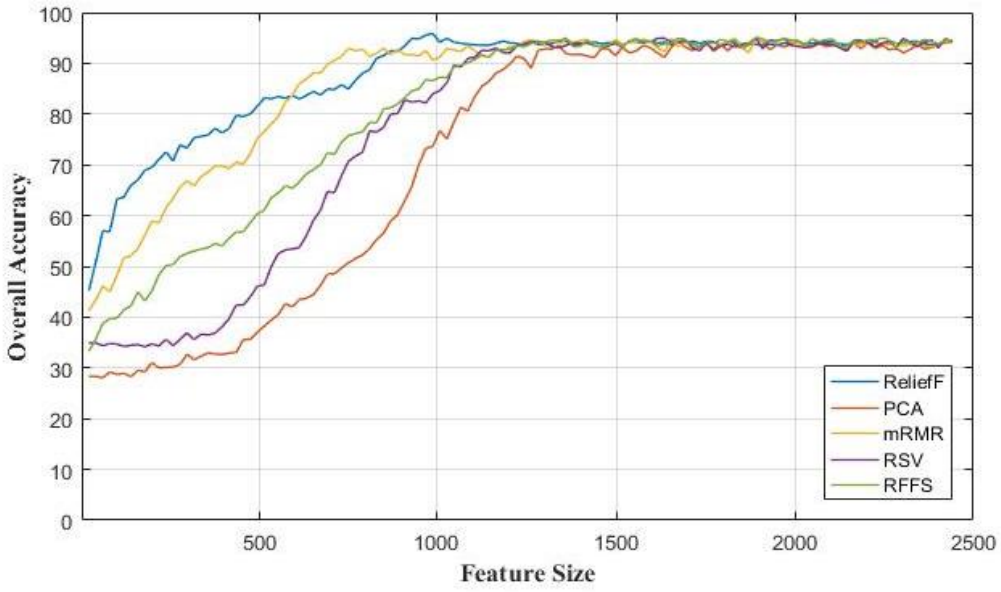


Figure 5-4 Improvement in overall accuracy rates by ANN classifier while increasing the number of the selected features and applying different feature selection algorithms on rabbit RGB CRF

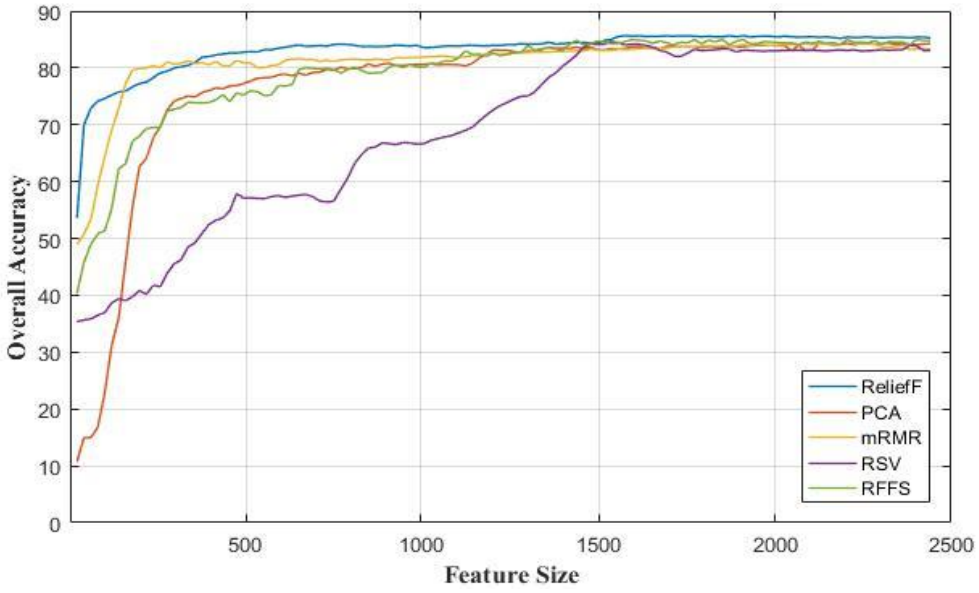


Figure 5-5 Improvement in overall accuracy rates by K-NN classifier while increasing the number of the selected features and applying different feature selection algorithms on rabbit RGB CRF

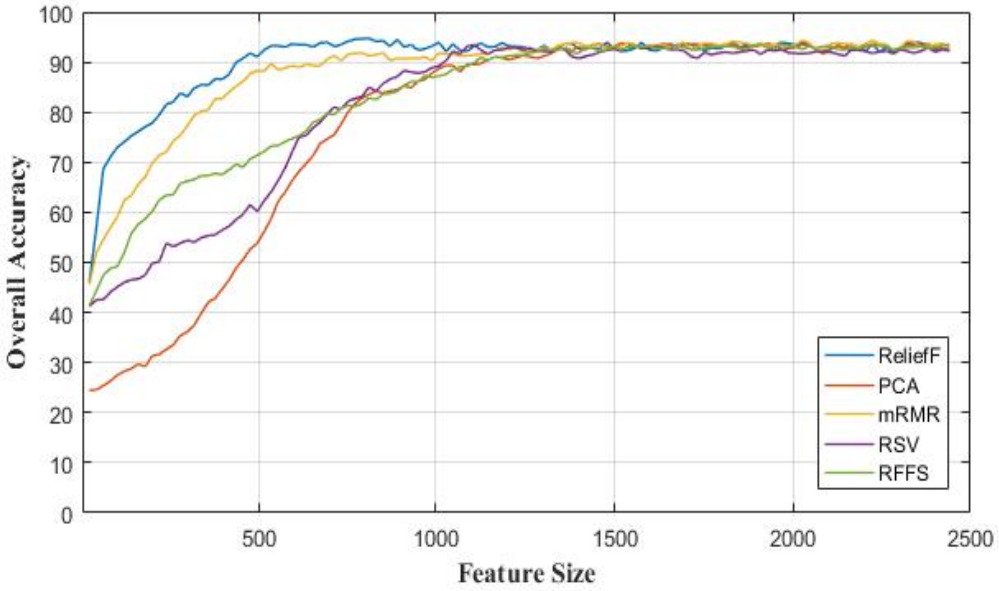


Figure 5-6 Improvement in overall accuracy rates by SVM classifier while increasing the number of the selected features and applying different feature selection algorithms on rabbit RGB CRF

5.3 Optimizing Results from RGB CRF in the Minimized Images

As discussed in the previous chapter, applying image resizing methods to minimize the image matrices by up to one-quarter did not cause misclassification problems. Therefore, applying the feature selection methods on this new database of smaller images might present a space with smaller feature dimensions while retaining high accuracy rates. This section presents the results from the one-quarter size datasets from both image databases using nearest neighbour interpolation (NNI). The reason of choosing NNI is due to the fact that it is the easiest interpolation method in term of the implementation time and the computational simplicity. Additionally, in Chapter 4, the results obtained based on minimizing the sizes of images did not show major differences because of the type interpolation method applied, in contrast, most of the results were identical.

5.3.1 Results from RGB CRF of Minimized Chicken Images

The database of images of chicken Eimeria species has been prepared at the image processing stage with image matrix sizes of 510-by-370 pixels. Minimizing this size to one-quarter gives 128-by-93 pixels shown in the previous chapter. The results from the classification of optimizing pixel-based features using the selection and reduction algorithms applied in this research are presented below.

The selection of the most relevant features from this dataset, which contains 663 features, shows excellent overall accuracy rates given the small feature space found using the feature selection and reduction techniques considered in this research. Figures 5-7 - 5-9 display the progress curves of overall accuracy rates with regard to increasing the feature numbers. They show that the results when using ReliefF and mRMR with the three classifiers always have high starting rates as well as steep improvements feature number in comparison with the other selection methods. 96.79% (±2.54%) was the best accuracy rate found, which obtained using ANN with 439 features from ReliefF selection method. Constituting 66% of this dataset's feature space. Additionally, ANN with ReliefF achieved 92.15% accuracy when only 70 features were processed. This number of features represents only 10% of the entire size of this dataset. SVM also yielded very good results. Its accuracy reached 94.80% (±2.33%) with the 138 most relevant features selected using ReliefF method. Moreover, it obtained 95.84% (±2.4%) as the maximum rate analysing 451 features selected by ReliefF. SVM with the other feature selection algorithms yielded various results, but accuracy rates always exceeded 93%. In regard to K-NN results, the best accuracy rate is 94.70% (±2.84%) from 180 features selected by ReliefF. The overall accuracy rate for K-NN reached 91% with only the 30 most

relevant features selected by ReliefF. Furthermore, K-NN results with mRMR, PCA, RSV, RFFS are less than those with ReliefF, but all accuracy rates are higher than 92.5%.

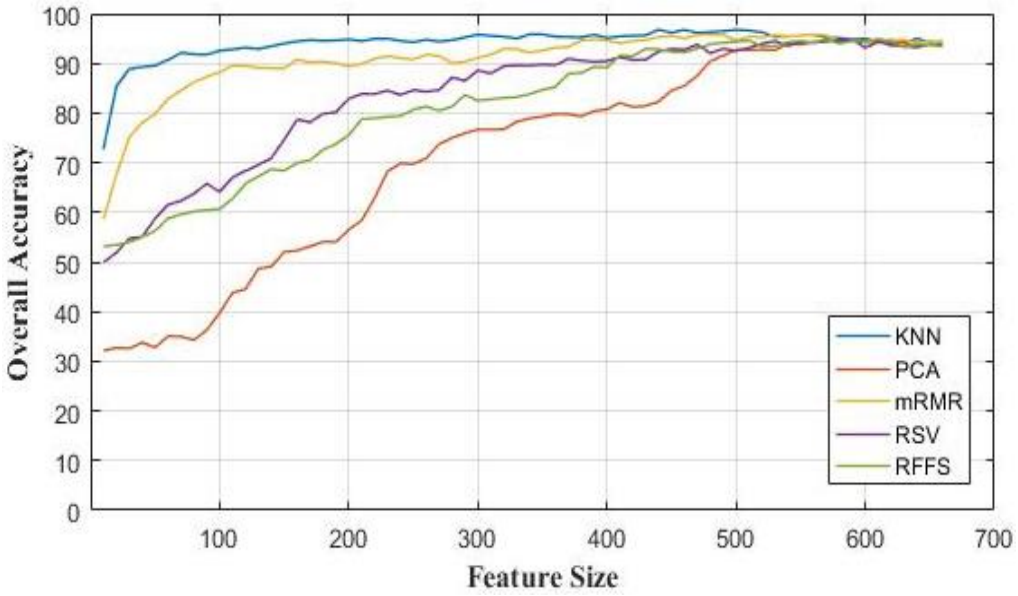


Figure 5-7 Improvement in overall accuracy rates using ANN classifier while increasing the number of selected features and applying different feature selection algorithms on the one-quarter image sizes of chicken RGB CRF

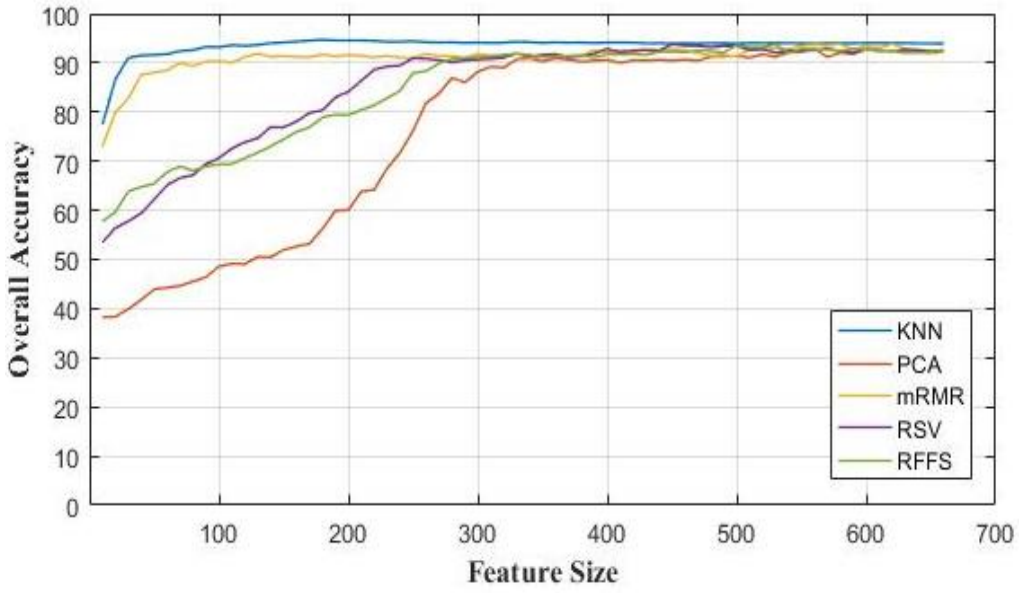


Figure 5-8 Improvement in overall accuracy rates using K-NN classifier while increasing the number of selected features and applying different feature selection algorithms on the one-quarter image sizes of chicken RGB CRF

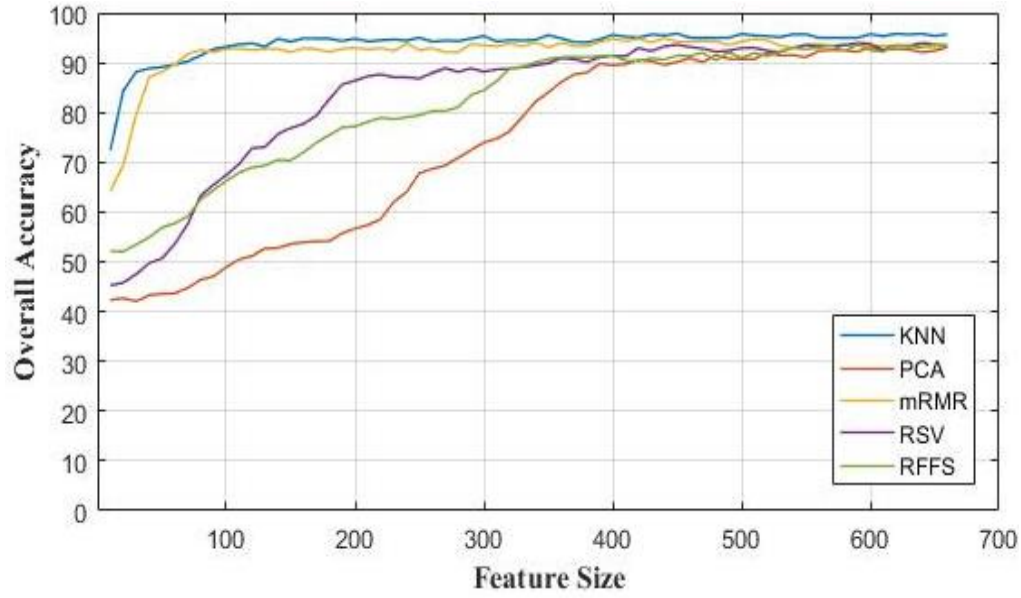


Figure 5-9 Improvement in overall accuracy rates using SVM classifier while increasing the number of selected features and applying different feature selection algorithms on the one-quarter image sizes of chicken RGB CRF

5.3.2 Results from RGB CRF Sets of Minimized Rabbit Images

The analysis of the microscopic images of the rabbit database showed various overall accuracy rates whether processed with the whole feature spaces or when selecting the most relevant variables using different methods. This section illustrates the results obtained based on the selection of the most relevant variables from the minimized images in the dataset. As explained in the previous chapter, the one-quarter image sizes of the rabbit database are 121-by-84 pixels.

RGB CRF includes 615 features. Selection of the most relevant ones leads to improvement in the accuracy rates, as shown in Figures 6-73 - 6-75. As can be noticed in these figures, the K-NN results did not reach 90% with any of the selection methods. K-NN results from this dataset is lower than RGB CF, which implies that the addition of RGB RF confuses K-NN when combined with RGB CF for this dataset. Both ANN and SVM curves reached more than 93% accuracy with all feature selection techniques. The best accuracy rates are 95.96% (±2.94%) from 403 features and 95.73% (±2.62%) from 362 features using ANN and SVM respectively with ReliefF algorithm.

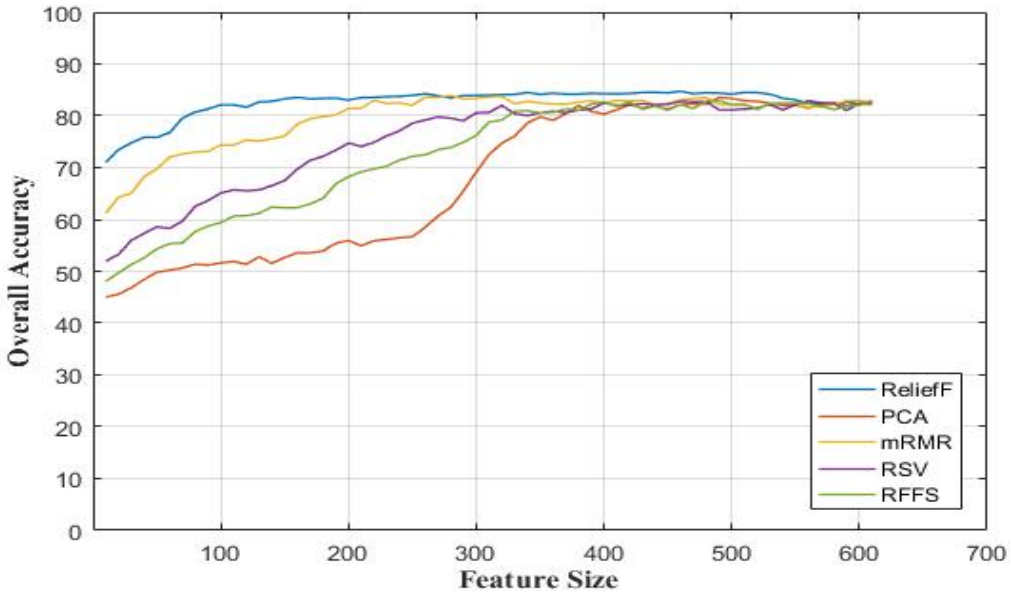


Figure 5-10 Improvement in overall accuracy rates using K-NN classifier while increasing the number of selected features and applying different feature selection algorithms on the one-quarter image sizes of rabbit RGB CRF

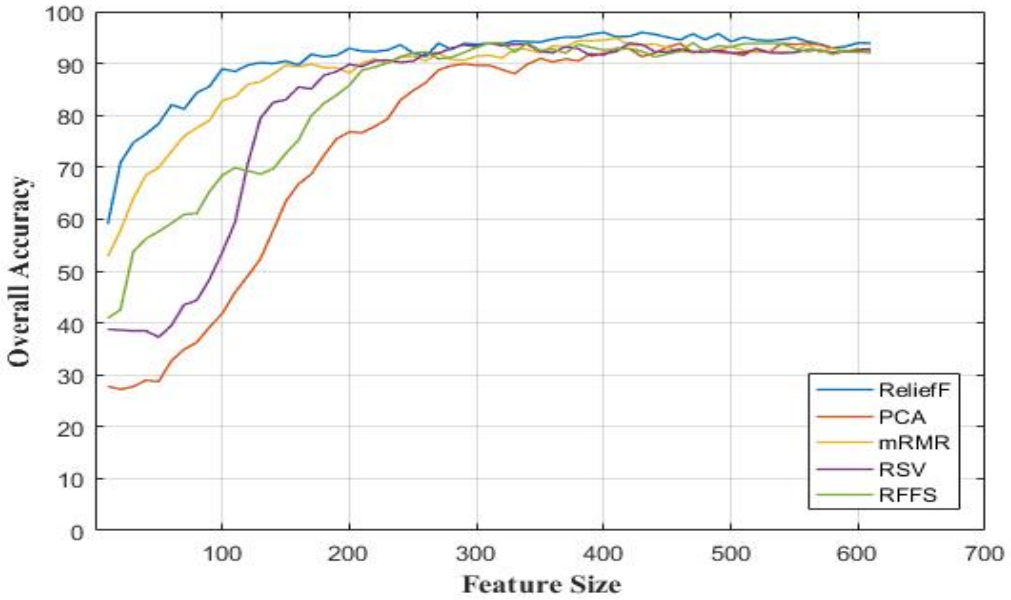


Figure 5-11 Improvement in overall accuracy rates using ANN classifier while increasing the number of selected features and applying different feature selection algorithms on the one-quarter image sizes of rabbit RGB CRF

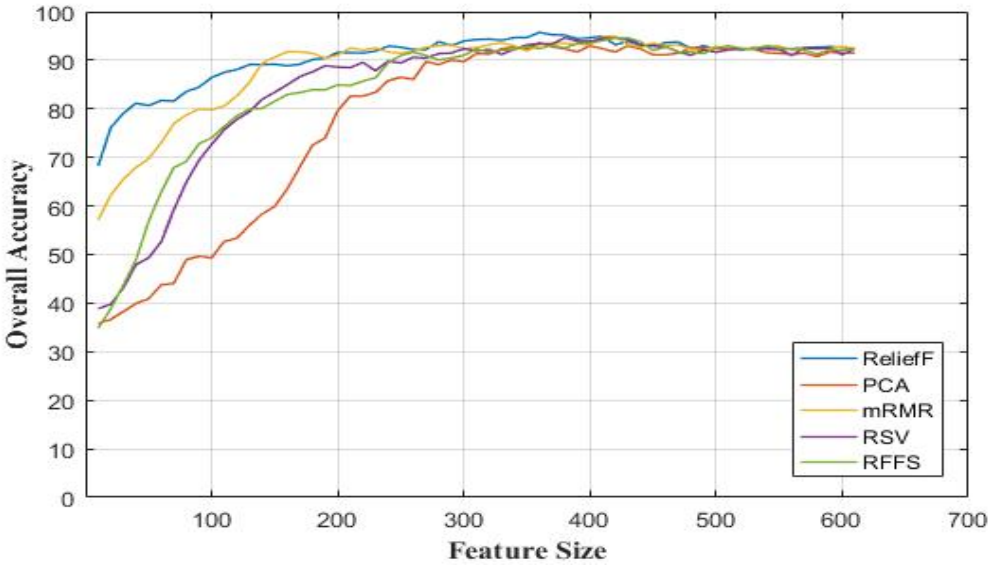


Figure 5-12 Improvement in overall accuracy rates using SVM classifier while increasing the number of selected features and applying different feature selection algorithms on the one-quarter image sizes of rabbit RGB CRF

5.4 Conclusions of the Results

The results discussed in this chapter are better than those described in the previous chapter in terms of the levels of overall accuracy and the new feature spaces obtained using the selection and reduction methods utilized in this research. In general, ReliefF succeeded in extracting the most relevant features which supported the three classifiers in yielding their best results with the smallest feature spaces compared with the other selection methods. Moreover, ANN often provided the best accuracy results for the majority of the feature sets analysed.

5.5 Summary

This chapter has discussed the ability of the optimization of pixel-based feature datasets to increase overall accuracy rates as well as to reduce the dimensions of feature sets. It could achieve these aims by reducing the feature space to much smaller sizes up to 40% of the original size of the dataset. Accuracy rates for both chicken and rabbit databases have been improved up to 96.70% and 95.85% respectively. The two highest accuracy rates were yielded using ANN. The standard deviations of results achieved for all datasets did not exceed ±3%, which means that the accuracy rates are quite reliable. This level dispersion was achieved for the selected features from datasets of the original sizes and those of one-quarter. For the feature selection methods considered, ReliefF succeeded in selecting the most useful features, followed by mRMR, reflected in their best results achieved with all classification models and datasets. PCA always ranks as the least effective techniques, and RSV and RFFS are in between the best accuracy rates and those from PCA. Although the results using K-NN classifiers are often the worst, the effectiveness of the classification models with the datasets analysed has shown reliable findings. The chapter next extracts super-pixel datasets based on ReliefF selection algorithm which always shows the ability to choose the most relevant features. The super-pixel datasets are extracted directly from the images without considering datasets of CF, RF, and CRF.

CHAPTER 6 DEVELOPMENT OF SUPER-PIXEL FEATURE SET

Extraction of the super-pixel sets is discussed in this chapter. In the experiments shown previously ReliefF method succeeded in the selection of the most relevant features from both image databases in greyscale level and RGB datasets which supported classifiers in most cases to achieve the best accuracy rates analysing those features selected by ReliefF. Therefore, ReliefF is applied to segment the super-pixel sets based on the values of pixels to identify the regions of interest instead of processing the features of CF, RF, and CRF. As described before, these CF, RF, and CRF are the means of the pixel values of columns, rows, and merging them together. In this chapter, ReliefF is applied on the pixels of images without any pre-segmentation. Therefore, every image in the database will be driven through ReliefF. Consequently, the features of the super-pixel feature set represent the real pixel values from every image. The super-pixel will be extracted from the images of the original and the one-quarter image sizes in greyscale and RGB for both databases. The classification results for the super-pixel feature sets are discussed below.

6.1 Analysis of Super-pixel Sets of the Chicken Database

The image-processing phase prepared dimensions the regions of interests of the chicken images in matrices of 510-by-370 pixels. So, the total number of pixels in every image is 188700 pixels for the greyscale images and 566100 pixels for colours. Additionally, the image sizes of one-quarter are 128-by-93 pixels. So, the number of pixels are 11904 and 35712 for the greyscale level and colours images respectively.

6.1.1 Selection of the Super-pixel Feature sets

The segmentation of super-pixel sets using ReliefF method detects the most relevant pixels which are allocated in the edges of the oocysts for both of the original or one-quarter image sizes. It also applies for greyscale level and colour images. Figure 6-1 shows samples of the species of E.Maxima extraction of various spaces of super-pixel feature sets from the one-quarter sizes of the colour images. The appearances of super-pixel sets for the original size images of Eimeria species in chickens are shown in Appendix E.

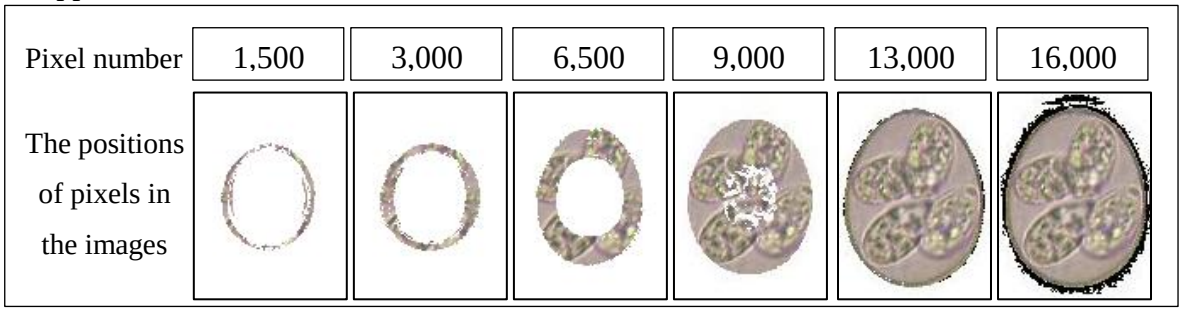


Figure 6-1 Samples of different spaces of the super-pixel feature sets and their positions in E. Maxima species of the one-quarter size images of chicken Eimeria

6.1.2 Results for Classification of Super-pixel Feature Sets

The results obtained for the super-pixel datasets of Eimeria species in the chicken database were all high for the original and the one-quarter size images. Table 6-1 shows results yielded for the grey super-pixels datasets of different dimensions using all classifiers. It can be noticed that ANN with a super-pixel set of 100 features achieved an accuracy rate of 80.09% (±2.55%). K-NN and SVM obtained lower accuracy rates when analysing the same dimensions of the feature set yielding 68.02% (±0.87%) and 74.21% (±1.97%) respectively. All of these results increased when the classifiers analysed 500 features of the super-pixel set. Figure 6-2 and 7-3 show the improvement in classification results for the grey super-pixels using all classifiers. The accuracy rates

acquired sharply increased when analysing very small feature spaces. ANN showed the best classification result of 84.42% (±2.29%) when processing 500 features. Furthermore, accuracy rates did not show any significant improvement when the feature space contained more 1000 features. In contrast, all results declined to different levels when the number of features exceeded 4000.

Table 6-1 Results for the grey super-pixel feature sets of the original images of the chicken database

Classifiers		Super-pixel Number				
		100	500	1,000	5,000	15,000
K-NN	μ	68.02%	77.25%	79.05%	79.19%	77.83%
	σ	0.87%	0.80%	0.78%	0.79%	0.49%
ANN	μ	80.09%	**84.42%**	84.20%	84.29%	82.39%
	σ	2.55%	2.29%	2.68%	2.45%	2.59%
SVM	μ	74.21%	81.49%	81.67%	81.38%	80.98%
	σ	1.97%	1.20%	1.62%	0.81%	0.93%

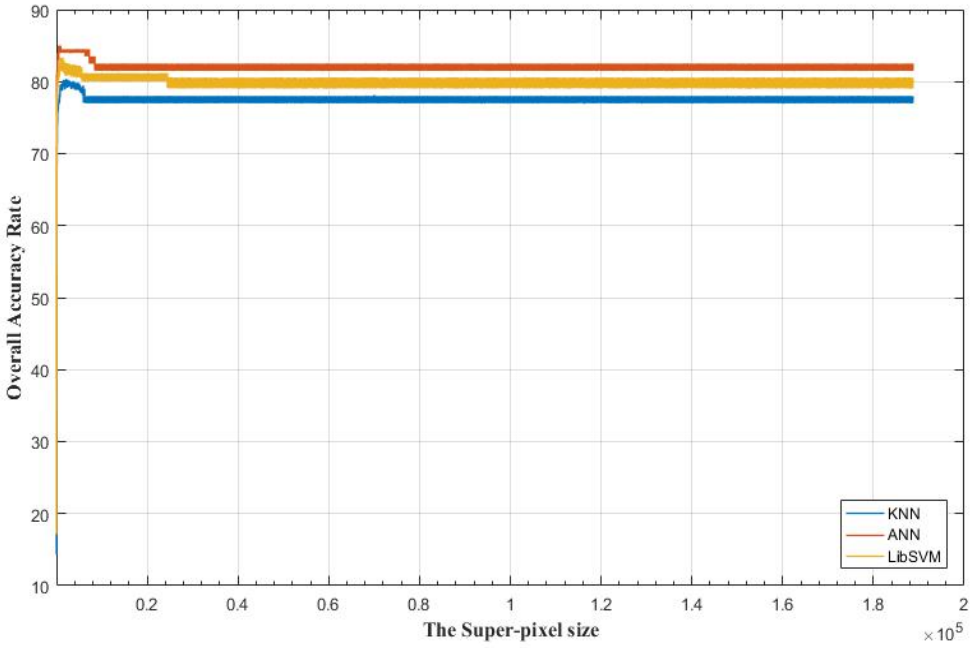


Figure 6-2 Overall accuracy rates obtaining from the original image sizes of chickens for the grey super-pixel feature sets using K-NN, ANN, and SVM

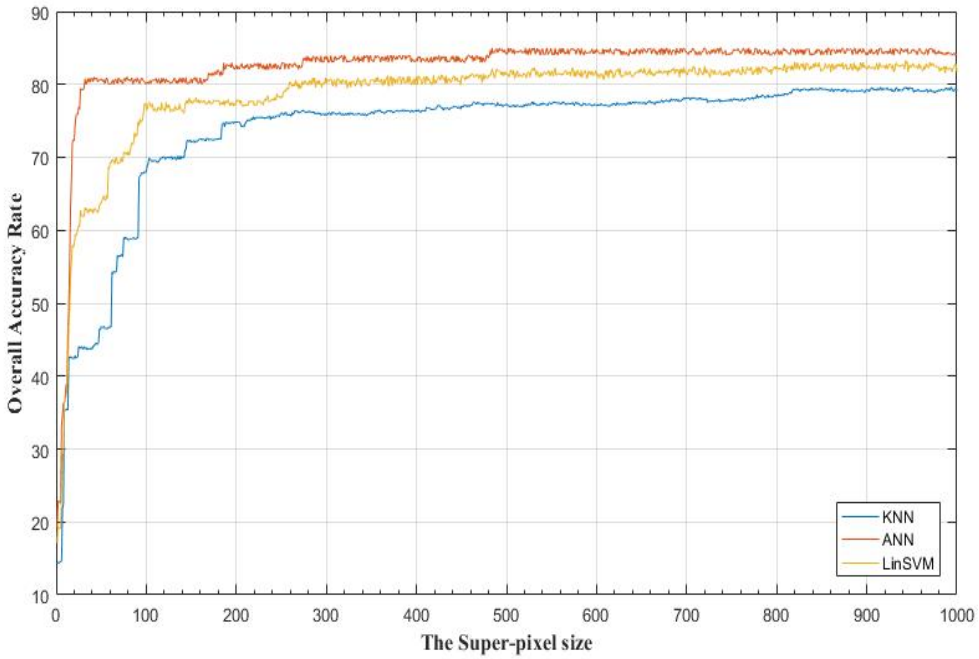


Figure 6-3 Overall accuracy rates for the grey super-pixels sets from chicken images of 1 to 1000 features using K-NN, ANN, SVM

Classification of the colour super-pixels gave results higher than those for the grey images, but with higher number of features. Table 6-2 shows samples of the accuracy rates obtained with different sizes of super-pixel set spaces. ANN still provides the best results compared with SVM and K-NN. The best accuracy rates for all super-pixel feature set was 93.54% (±1.68%) obtained from the super-pixel of 1800 features using ANN. The results yielded using ANN rapidly increased to achieve 91.63% (±2.06%) when processing 634 features. SVM provided the second best rates and the results of K-NN are the lowest. Their best accuracy rates were 89.97% (±1.8%) and 91.97% (±1.44%) obtained from 1080 and 1780 features respectively. shows the improvement in the accuracy rates while increasing the super-pixel spaces.

Table 6-2 Results for the colour super-pixel feature sets of the original image sizes of the chicken database

Classifiers		Super-pixel Number				
		634	1,800	6,000	10,000	20,000
K-NN	μ	83.99%	86.80%	82.65%	82.92%	83.27%
	σ	1.88%	1.49%	2.04%	1.67%	1.8%
ANN	μ	91.63%	**93.54%**	87.93%	87.16%	86.6%
	σ	2.06%	1.68%	1.88%	2.01%	2.42%
SVM	μ	83.97%	91.33%	86.75%	88.27%	85.11%
	σ	1.6%	1.56%	1.52%	1.89%	1.78%

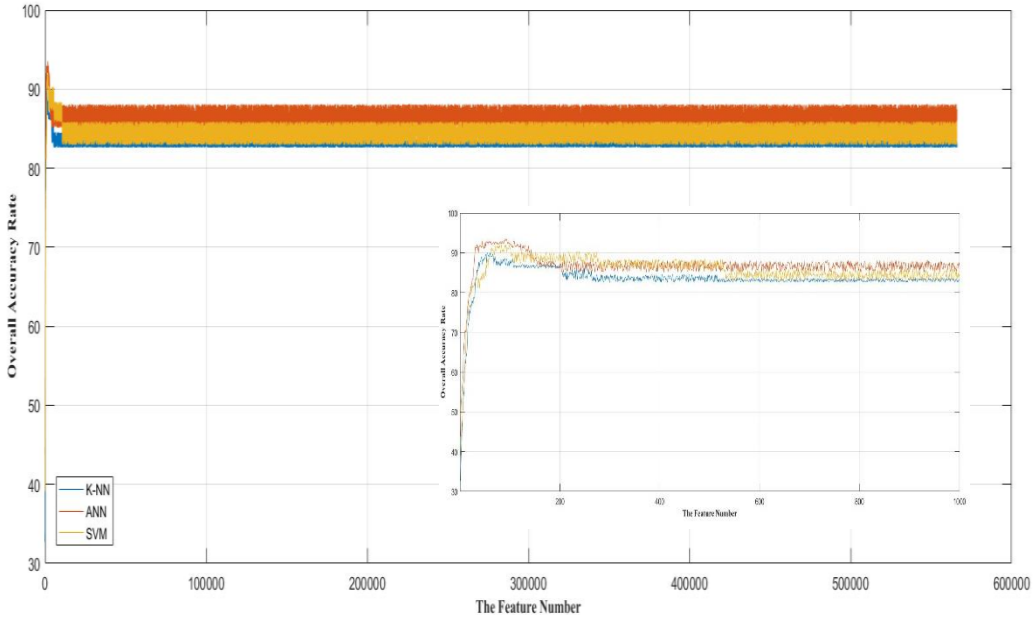


Figure 6-4 Overall accuracy rates obtaining from the original image sizes of chickens for the colour super-pixel feature sets using K-NN, ANN, and SVM

Furthermore, the classification for the super-pixel sets extracted from the one-quarter image sizes has also shown considerable accuracy rates for both the grey and colour datasets. ANN again achieved the best accuracy rates, obtaining 81.47% (±2.1%) when processing 544 features out of the total of 11904 of the grey features. Moreover, its best overall accuracy rate was 84.14% (±2.6%) when analysing 1124 features. SVM and K-

NN needed larger numbers of features to achieve their best accuracy rates, which were lower than those for ANN, as shown in Table 6-3 and Figure 6-5.

Table 6-3 Results for the grey super-pixel feature sets of the one-quarter size images of the chicken database

Classifiers		Super-pixel Number				
		540	1,124	6,000	10,000	11,904
K-NN	μ	68.43%	81.02%	76.19%	76.3%	75.96%
	σ	1.78%	1.64%	1.33%	2.8%	2.89%
ANN	μ	81.47%	**84.14%**	81.62%	81.24%	81.56%
	σ	2.1%	2.6%	2.7%	2.35%	2.47%
SVM	μ	78.95%	82.47%	77.97%	77.17%	77.65%
	σ	1.22%	1.08%	1.67%	1.87%	2.47%

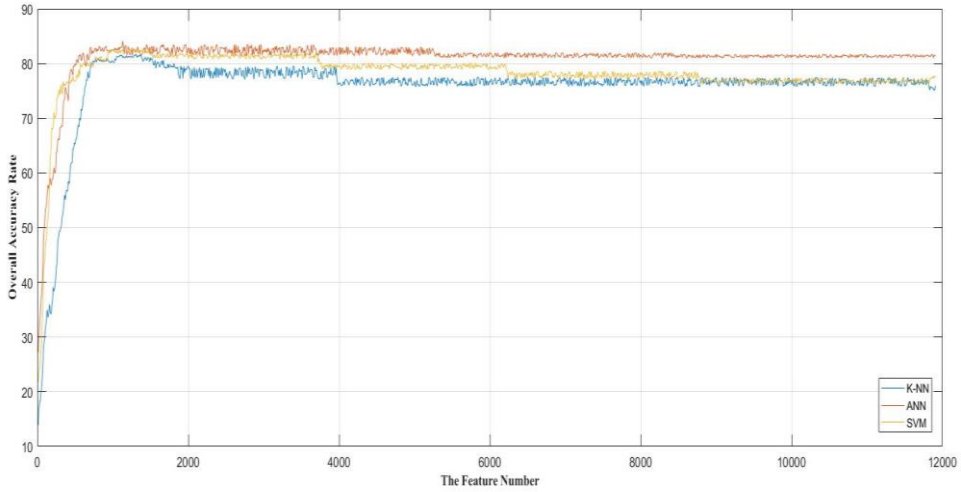


Figure 6-5 Overall accuracy rates obtaining from the one-quarter image sizes of chickens for the grey super-pixel feature sets using K-NN, ANN, and SVM

For the super-pixel sets segmented from the colour images of one-quarter size, the best accuracy rate was better than that for the grey datasets, as shown in Table 6-4 . A performance of 93.96% (±2.03%) was yielded from processing of 867 features using ANN. Figure 6-6 shows the improvement in the accuracy rates yielded using ANN, K-NN, and SVM. It can be noticed that the results for all classifiers improved when the size of the feature space was smaller than 2000. In addition, the accuracy rates for all

classification models decreased slightly when processing feature sets of more than 4000 features. However, results for SVM with bigger number of the super-pixels remained the highest, but did not reach results for ANN. The best accuracy rates yielded for K-NN and SVM were 88.74% (±2.4%) from 923 features and 90.75% (±1.63%) from 702 features respectively.

Table 6-4 Results for the colour super-pixel feature sets of the one-quarter size images of the chicken database

Classifiers		Super-pixel Number				
		867	2,000	6,000	10,000	20,000
K-NN	μ	87.63%	86.66%	85.05%	83.32%	78.21%
	σ	1.8	2.43%	1.62%	2.05%	2.61%
ANN	μ	**93.96%**	91.76%	90.01%	86.14%	80.92%
	σ	2.03	2.41%	2.64%	2.36%	2.84%
SVM	μ	89.91%	88.9%	87.74%	89.07%	87.04%
	σ	1.75	2%	1.98%	2.17%	2.36%

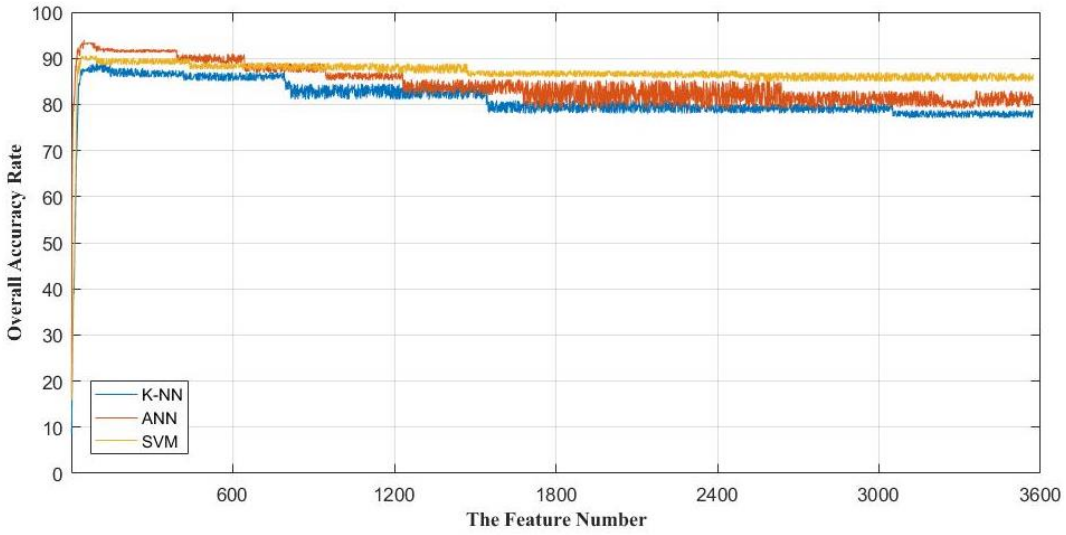


Figure 6-6 Overall accuracy rates obtaining from the one-quarter image sizes of chickens for the colour super-pixel feature sets using K-NN, ANN, and SVM

6.2 Analysis of the Super-pixel Sets of the Rabbit Database

The dimensions of the original images of the rabbit database were prepared based on the image processing stage in this research in matrices of 482-by-335 pixels. So, the total number of pixels in every grey image is 161470 and the colour images 484410 pixels. Moreover, the dimensions of the one-quarter size images are 121-by-84 pixels. The of numbers pixels in the one-quarter size images are 10164 and 30492 pixels for grey and colour images respectively.

6.2.1 Selection of the Super-Pixel Feature Sets

The most relevant features for the super-pixel feature set extracted using ReliefF are allocated along the edges of the oocysts for both sizes of images. Figure 6-7 shows samples of E.Magna with different sizes of feature spaces. It shows that ReliefF selects from both databases the most relevant pixels those at the boundaries of oocysts. Then the relevant pixels are extended inside the oocysts. For the small oocysts the most relevant pixels are zeros, because they are for the background as shown in Appendix E. This might affect the classification outcomes when processing small sizes of feature spaces.

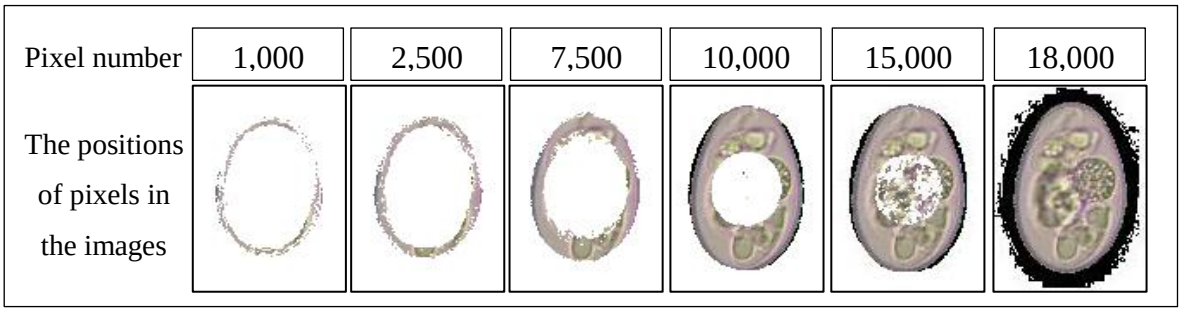


Figure 6-7 Samples of different spaces of the super-pixel feature sets and their positions in E. Magna species of the one-quarter size images of rabbit Eimeria

6.2.2 Results for Classification of the Super-pixel Feature Sets

The classification results for the super-pixel sets of the rabbit images are still lower than those for the chicken images. ANN always achieves the best accuracy rates compared with SVM and K-NN. Furthermore, the accuracy rates for colour feature sets are better than those for grey sets. Table 6-5 and 7-6 show the accuracy rates yielded for various sizes of grey and colour super-pixel datasets using the three classifiers and Figure 6-8 and 7-9 show the improvement in the accuracy rates obtaining for the grey and colour super-pixel sets. The best accuracy rates of 81.65% (±1.83%) and 90.70% (±2.21%) were obtained by ANN when processing 484 and 1083 for grey and colour features respectively.

Table 6-5 Results for the grey super-pixel feature sets of the original size images of the rabbit database

Classifiers		Super-pixel Number				
		484	1000	6,000	10,000	16,000
K-NN	μ	68.30%	79.20%	69.77%	69.74%	69.75%
	σ	1.84%	2.35%	2.65%	2.38%	2.3%
ANN	μ	**81.65%**	74.85%	71.76%	73.48%	71.61%
	σ	1.83%	2.68%	2.47%	2.36%	2.67%
SVM	μ	80.01%	77.63%	76.03%	74.71%	75.64%
	σ	2.04%	1.93%	2%	2.08%	2.11%

Table 6-6 Results for the colour super-pixel feature sets of the original size images of the rabbit database

Classifiers		Super-pixel Number				
		1,083	3,000	6,000	10,000	25,000
K-NN	μ	86.85%	86.14%	80.10%	77.04%	77.95%
	σ	2.44%	2.65%	2.77%	2.3%	2.52%
ANN	μ	**90.70%**	86.97%	86.51%	86.54%	85.13%
	σ	2.21%	2.47%	2.5%	2.66%	2.72%
SVM	μ	87.81%	87.15%	85.83%	82.23%	83.03%
	σ	1.98%	2.3%	1.69%	2.04%	2.4%

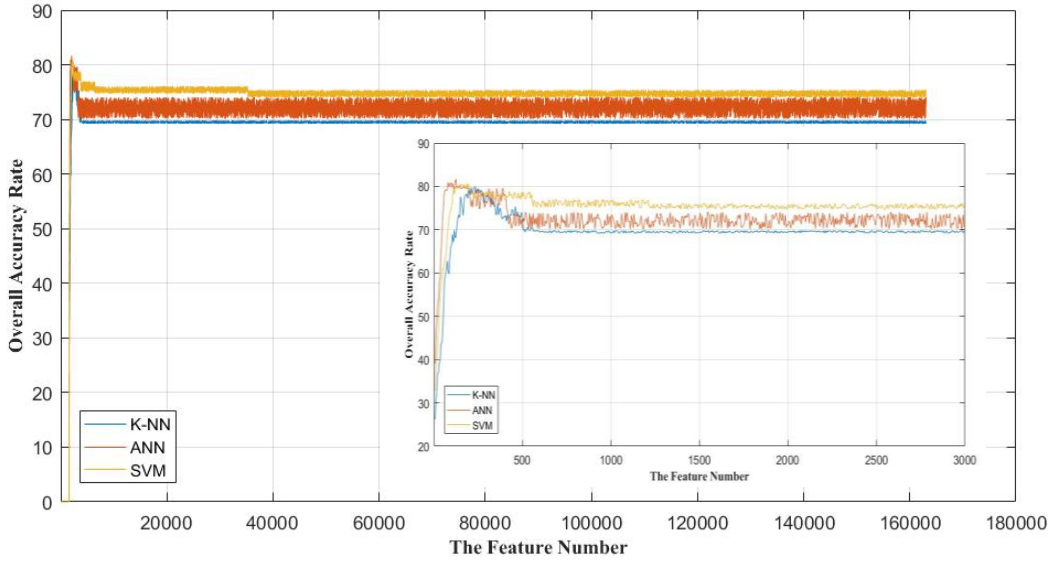


Figure 6-8 Overall accuracy rates obtaining from the original image sizes of rabbits for the grey super-pixel feature sets using K-NN, ANN, and SVM

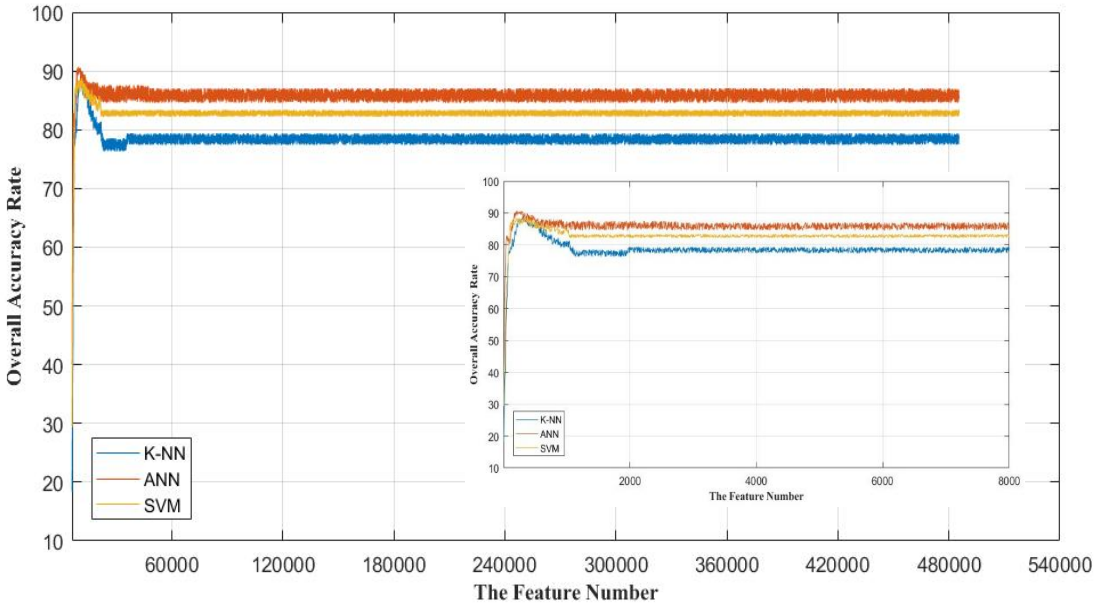


Figure 6-9 Overall accuracy rates obtaining from the original image sizes of rabbits for the colour super-pixel feature sets using K-NN, ANN, and SVM

The classification of the one-quarter size images has also yielded good results in terms of high accuracy rates and the small feature spaces. Table 6-7 and 7-8 show the highest overall accuracy rates for several sizes of super-pixel sets of the grey and colour one-quarter size images. The classification of the 743 most relevant grey features using ANN achieved 80.05% (±2.3%) accuracy. The best accuracy rate for the grey super-pixels was 80.47% (±1.93%) using ANN when processing the 1460 most relevant features. SVM also yielded good results when analysing the grey super-pixel dataset. It achieved 80.2% (±1.8%) accuracy when processing the 2763 most relevant features.

For the classification of the colour super-pixel sets, ANN also gave the best accuracy rate when analysing the smallest feature space compared with K-NN and SVM. ANN obtained 89.86% (±2.98%) as the overall best accuracy rate when processing the 3246 most relevant features. The highest accuracy rates obtained using K-NN and SVM were 84.32% (±2.28%) from 2309 features and 88.64% (±1.95%) from 972 features respectively. Figure 7-6 and 7-7 depict the improvements in accuracy rates for the grey and colour super-pixel sets while the dimensions of the feature spaces increased.

Table 6-7 Results for the grey super-pixel feature sets of the one-quarter size images of the rabbit database

Classifiers		Super-pixel Number				
		743	1460	2463	5,000	10,000
K-NN	μ	62.20%	74.50%	77.62%	74.72%	75.00%
	σ	1.77%	2.43%	1.8%	1.73%	2.04%
ANN	μ	80.05%	**80.47%**	79.83%	79.23%	78.47%
	σ	2.3%	1.93%	2.36%	2.8%	2.68%
SVM	μ	66.79%	76.41%	79.92%	76.02%	73.60%
	σ	1.68%	2.14%	2.64%	2.04%	2.74%

Table 6-8 Results for the colour super-pixel feature sets of the one-quarter size images of the rabbit database

Classifiers		Super-pixel Number				
		972	2,309	3,246	7,000	20,000
K-NN	μ	77.68%	84.32%	79.62%	79.72%	79.45%
	σ	2.3%	2.28%	2.53%	2.71%	2.3%
ANN	μ	84.03%	87.61%	**89.86%**	85.53%	85.38%
	σ	2.85%	2.44%	2.98%	2.73%	2.7%
SVM	μ	88.64%	87%	85.19%	83.65%	83.74%
	σ	1.95%	2.09%	2.32%	2.47%	2.18%

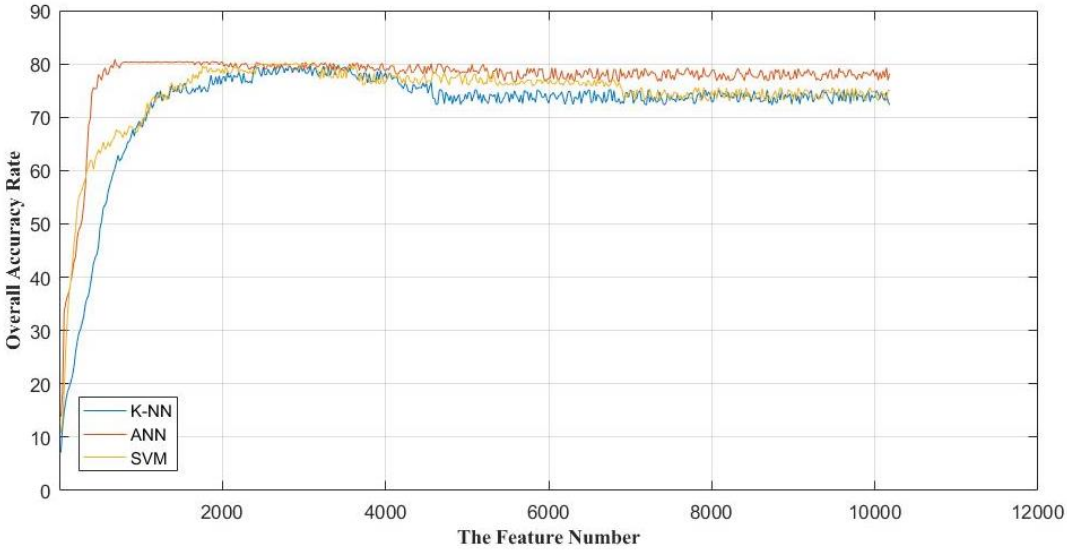


Figure 6-10 Overall accuracy rates obtaining from the one-quarter image sizes of rabbits for the grey super-pixel feature sets using K-NN, ANN, and SVM

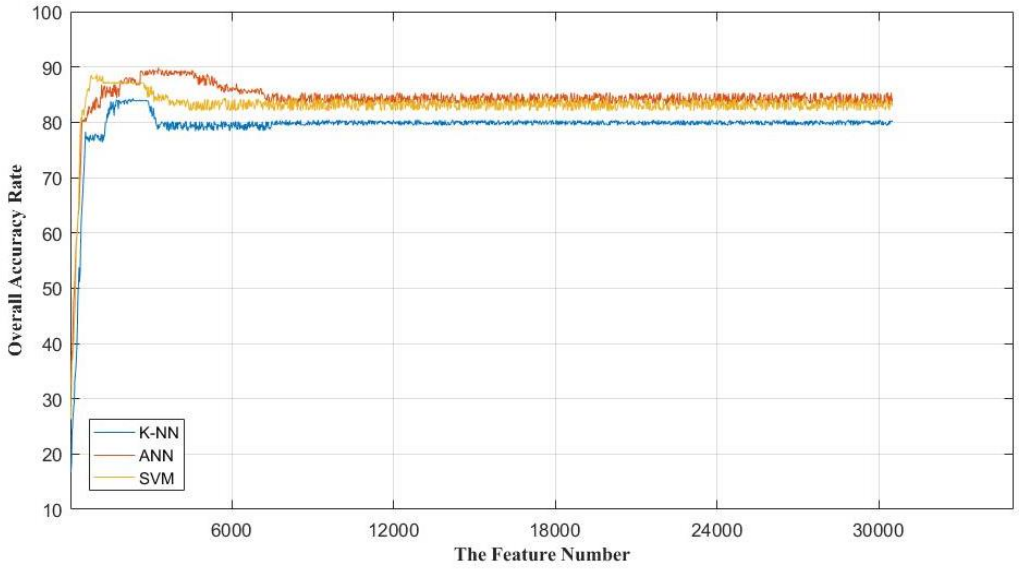


Figure 6-11 Overall accuracy rates obtaining from the one-quarter image sizes of rabbits for the colour super-pixel feature sets using K-NN, ANN, and SVM

6.3 **Summary**

This chapter has discussed the results obtained for the super-pixel feature sets in both databases of grey and colour images. The best accuracy rates for the super-pixel feature sets were lower than those for RGB CRF presented in the previous two chapters. However, the extraction of the super-pixel sets did not require any calculation, whereas in contrast CF, RF, and CRF sets require the mean values of the columns and rows of the oocyst images to be calculated. This chapter reported that super-pixel sets are extracted using ReliefF method from the pixels themselves. Although the accuracy rates were lower than for RGB CRF, they exceeded 90% in several cases especially for the colour feature sets when small feature spaces were processed. The classification of super-pixel feature sets of the chicken images gave results higher than those for the rabbit images. Furthermore, ANN still provided the best accuracy rates for both chicken and rabbit images. Finally, this chapter has shown that the super-pixel feature sets can be used as informative descriptors to distinguish Eimeria species.

CHAPTER 7 CONCLUSIONS AND FUTURE WORK

7.1 Contributions of the thesis

Since the last century, artificial intelligence (AI) techniques have been utilized in different areas to provide automated and convenient solutions for many problems in life. Medical issues constitute one of the most important challenges which have been explored using various machine learning methods. The recognition of medical images based on image processing and machine learning techniques have been investigated in this thesis. This research has proposed pixel-based features to identify Eimeria species from microscopic images. Eimeria is one of the single-celled intestinal parasites which infects animals and leads to death in a very short time if not diagnosed and treated. This issue has been previously considered by analysing morphological characteristics. These features cannot always be powerful descriptors for discrimination, however because of the close similarities between microscopic objects. Furthermore, the used for extraction processes of such features required are very complex. Therefore, this research has proposed features based on the pixels in the greyscale and colour images. The analysis of these features sows that they are very informative in identifying Eimeria species from microscopic images. The motivation of this research is to propose a robust identification method and to find alternative features instead of morphological ones.

Chapter 2 has summarised existing automated studies which have tried to provide reliable identification methods, exploring the features utilized to distinguish regions of interest, the extraction and classification processes used, and the results of

previous studies. Additionally, it sought strengths and weaknesses in automated studies, which show that many different morphological characteristics have already been exploited as descriptors to identify different objects. In addition, the classification tasks in these studies were considered using several models. Some studies extract very complicated features; nevertheless, they often achieved very poor results.

Chapter 3 explained the methodology proposed to classify Eimeria species built on the use of pixel-based features. The steps involved started from an image processing stage to detect oocysts of Eimeria in images and then to allocate them according to the proposed identification conditions. This is followed by the extraction of the desired features which are column features (CF), row features (RF), and a combination of column and row features (CRF). These feature sets are extracted from greyscale and colour images and from every image channel (red, green, and blue). Moreover, the extraction of the super-pixel datasets. All of these feature sets are optimized using feature ReliefF, mRMR, and PCA selection and reduction techniques. This research also proposed RSV and RFFS as selection methods. Furthermore, the proposed method utilizes three classification algorithms to process the features extracted: K-NN, ANN, and SVM. Finally, the steps taken in the evaluation of the classification findings are described.

In Chapter 4, the results of the implementation of the proposed research methodology were discussed. These results are highly reliable for both image databases. The results obtained reached accuracy rates of 95.57% (±0.79%) and 94.21% (±1.79%) for the images of chickens and rabbits respectively, using ANN. The other two classifiers also obtained good accuracy levels, but lower than those for ANN. Furthermore,

minimizing image sizes up to one-third of the original size is shown to maintain high accuracy as does reducing the size of feature space. However, it is also found that the features extracted from separate image channels cannot be informative or provide satisfactory levels of accuracy.

Chapter 5 discussed the optimization of the feature number of RGB CRF which provided the best results in CH 4. The feature selection and reduction methods used could reduce number of the features up to 20% in some cases and improves the accuracy. The best results achieved were 96.7% (±0.89%) and 95.85% (±2.4%) from 21% and 40% of the total features in RGB CRF for the chicken and rabbit images respectively. The results are also high for all of the classifiers and selection methods considered when optimizing the feature sets of the one-quarter size images. The best results for those images were 96.79% (±2.54%) and 95.96% (±2.94%) for RGB CRF of the chicken and rabbit respectively.

As described in Chapter 6, the classification of the super-pixel feature sets showed that the analysis of the pixels extracted using ReliefF can be the most meaningful characteristics to recognize regions of interests. The high accuracy results achieved were obtained when processing various dimension of the super-pixels feature sets. In some cases good accuracy rates were obtained when considering small size of feature space. For example, ANN achieved 91.63% (±2.06%) when processing the 634 most relevant features of the colour super-pixel feature set of chickens. The classification of rabbit super-pixels yielded results lower than those for the chicken images. However, the results for rabbit grey and colour super-pixel sets were good.

In Conclusion, the discussion and the results shown in the chapters of this research

7.2 The research limitations

The main aim of this research is to find a simple, uncomplicated process that can be used to extract the proposed features This process will also be invaluable when classifying different Eimeria species. Despite the high overall accuracy rate yielded, in a few cases intervention is needed to avoid detection of the noise in the images. This requires finding more intensive algorithms to solve this problem.

7.3 Future Work

Despite the impressive results obtained in this research, more studies are required to adopt for the methodology proposed in this research to be made a more robust technique to identify the cases discussed or others. Future work is needed to extend the ideas covered and explained in this research as follows:

- Investigation of more datasets which include Eimeria species, other single-celled objects or different microscopic objects will confirm the hypothesis of the present research to identify similar to these objects analysing the pixel-based features instead of the morphological features. The analysis of other datasets which might include images with various levels of resolution will prove the qualities of the pixel-based features as meaningful descriptors to differentiate between microscopic objects. Additionally, this research has found that the separate analysis of colour channels images is not effective in identifying regions of interests. However, the analysis of other databases could show whether or not they are separately informative.

- The utilization of other classifiers might also achieve better results, as may exploiting different selection and reduction methods. This research suggests that the results obtained using ANN with ReliefF feature selection are always the best. Nonetheless, applying further classification models and selection algorithms may increase accuracy rates and reduce the size of the feature spaces. In addition, the K-NN and SVM classification models might yield higher results with different selection or reduction methods. This research applied filtered selection and reduction methods, whereas embedded or wrapper methods might be more effective in finding smaller feature set dimensions.
- The extraction of the super-pixel feature sets in this work was based only on the ReliefF method. Other techniques might be useful to select a fewer number of super-pixels and improve the accuracy rates. Furthermore, considering non-complex morphological characteristics beside the super-pixels could also improve the accuracy. In related work in this field, the extraction of super-pixel sets was previously done based on colour and position. This research considered the pixel values, but applying the other concepts as extra options might allow more informative super-pixel datasets.

REFERENCES

Abbas, R. Z., Iqbal, Z., Blake, D., Khan, M. N. and Saleemi, M. K. (2011) 'Anticoccidial drug resistance in fowl coccidia: the state of play revisited', *World's Poultry Science Journal*, 67(June 2011), pp. 337–350. doi: 10.1017/S004393391100033X.

Abdulla, S. U., Jayesh, H. T. G. and Nair, V. V. (2015) 'A General Approach for Color Feature Extraction of Microorganisms in Urine Smear Images', *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)*, pp. 338–341. doi: 10.1109/ICACC.2015.44.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and Süsstrunk, S. (2011) 'SLIC Superpixels Compared to State-of-the-art Superpixel Methods', *JOURNAL OF LATEX CLASS FILE FILES*, 6(1).

Acharya, T. and Ray, A. K. (2005) *Image Processing: Principles and Applications*. 4 Oct. 199*, A Jhon Wiley & Sons, Inc., Pubilcation*. 4 Oct. 199. New York, NY, USA: John Wiley \& Sons, Inc. doi: 10.1117/1.2348895.

Akbar, S., Jordan, L., Thompson, A. M. and Mckenna, S. J. (2015) 'Tumor Localization in Tissue Microarrays Using Rotation Invariant Superpixel Pyramids', in *2015 International Symposium on Biomedical Imaging (ISBI 2015)*.

Ali, R., Hussain, A. and Man, M. (2015) 'Feature extraction and classification for multiple species of Gyrodactylus ectoparasite', *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 13(3), pp. 503–511. doi: 10.11591/telkomnika.v13i3.7096.

Amami, R., Ben Ayed, D. and Ellouze, N. (2015) 'Practical Selection of SVM Supervised Parameters with Different Feature Representations for Vowel Recognition', *International Journal of Digital Content Technology and its Applications(JDCTA)*, 7(9), pp. 418–424. doi: 10.4156/jdcta.vol7.issue9.50.

Arco, J. E., Górriz, J. M., Ramírez, J., Álvarez, I. and Puntonet, C. G. (2015) 'Digital image analysis for automatic enumeration of malaria parasites using morphological operations', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2014.11.037.

Avci, D. and Varol, A. (2009) 'An expert diagnosis system for classification of human parasite eggs based on multi-class SVM', *Expert Systems with Applications*. Elsevier Ltd, 36(1), pp. 43–48. doi: 10.1016/j.eswa.2007.09.012.

Becker, E. R. (1976) *Coccidia and Coccidiosis of Domesticated, Game and Laboratory Animals and of Man, Veterinary Parasitology*. COLLEGIATE PRESS, INC. doi: 10.1016/0304-4017(76)90103-5.

Bolon-Canedo, V., Sanchez-Marono, N. and Alonso-Betanzos, A. (2015) 'Recent advances and emerging challenges of feature selection in the context of big data', *Knowledge-Based Systems*, 86, pp. 33–45. doi: 10.1016/j.knosys.2015.05.014.

Brown Jordan, A., Blake, D., Beard, J., Beharry, A., Serrette, L., Soleyn, A., Sookhoo, J., Blake, L., Brown, G. and Oura, C. (2018) 'Molecular Identification of Eimeria Species in Broiler Chickens in Trinidad, West Indies', *Veterinary Sciences*, 5(1), p. 12. doi: 10.3390/vetsci5010012.

Castañón, C. A. B., Fraga, J. S., Fernandez, S., Gruber, A. and Costa Luciano da F. (2007) 'Biological shape characterization for automatic image recognition and diagnosis of protozoan parasites of the genus Eimeria', *Pattern Recognition*, 40(7),

pp. 1899–1910. doi: 10.1016/j.patcog.2006.12.006.
Castañón B, C. A. B., Fraga, J. S., Fernandez, S., Pakandl, M., Costa, L. da F. and Gruber, A. (2007) *The Eimeria Image Database*. Available at: http://www.coccidia.icb.usp.br/imagedb/ (Accessed: 9 December 2014).
Chang, C. and Lin, C. (2013) 'LIBSVM : A Library for Support Vector Machines', *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, pp. 1–39. doi: 10.1145/1961189.1961199.
Chomboon, K., Chujai, P., Teerarassammee, P., Kerdprasop, K. and Kerdprasop, N. (2015) 'An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm', in *The Proceedings of the 2nd International Conference on Industrial Application Engineering 2015*, pp. 280–285. doi: 10.12792/iciae2015.051.
Cong, Y., Wang, S., Liu, J., Cao, J., Yang, Y. and Luo, J. (2015) 'Deep sparse feature selection for computer aided endoscopy diagnosis', *Pattern Recognition*. doi: 10.1016/j.patcog.2014.09.010.
Costa, L. da F. Da and Cesar-Jr., R. M. (2001) *Shape Analysis and Classification: Theory and Practice*. 1st edn. Boca Raton, FL, USA: CRC Press, Inc.
Daugschies, A., Imarom, S. and Bollwahn, W. (1999) 'Differentiation of porcine Eimeria spp. by morphologic algorithms', *Veterinary Parasitology*. doi: 10.1016/S0304-4017(98)00246-5.
Dogantekin, E., Yilmaz, M., Dogantekin, A., Avci, E. and Sengur, A. (2008) 'A robust technique based on invariant moments - ANFIS for recognition of human parasite eggs in microscopic images', *Expert Systems with Applications*, 35(3), pp. 728–738. doi: 10.1016/j.eswa.2007.07.020.
Duda, R. O., Hart, P. E. and Stork, D. G. (2000) *Pattern Classification*. 2nd edn. Wiley-Interscience.
Durgaba, R. P. L. (2014) 'Feature Selection using ReliefF Algorithm', *International Journal of Advanced Research in Computer and Communication Engineering*, 3(10), pp. 8215–8218.
Ertekin, S. (2009) *Learning in Extreme Conditions: Online and Active Learning with Massive, Imbalanced and Noisy Data*, *The Pennsylvania State University*. doi: 10.1017/CBO9781107415324.004.
Flores-Quispe, R., Velazco-Paredes, Y., Patino, R. E. E. and Castanon, C. A. B. (2014) 'Classification of human parasite eggs based on enhanced multitexton histogram', in *2014 IEEE Colombian Conference on Communications and Computing, COLCOM*. doi: 10.1109/ColComCon.2014.6860419.
Gardiner, C. H., Payer, R. and Dubey, J. P. (1988) *An atlas of protozoan parasites in animal tissues*, *United States Department of Agriculture*. Washington: U.S. Government Printing Office, Washington. doi: 10.1016/0169-4758(91)90176-O.
Gillespie, S. H. and Bamford, K. B. (2013) *Medical Microbiology and Infection at a Glance*. 4th edn. London: Wiley-Blackwell. doi: 10.1017/CBO9781107415324.004.
Han, C.-M., Lee, S.-W., Han, S. and Park, J.-S. (2011) 'Two-stage fish disease diagnosis system based on clinical signs and microscopic images', *Computational Science and Its Applications*, pp. 635–647. doi: 10.1007/3-540-68339-9_34.
Hira, Z. M. and Gillies, D. F. (2015) 'A review of feature selection and feature extraction methods applied on microarray data', *Hindawi Publishing Corporation Advances in*

Bioinformatics. doi: 10.1155/2015/198363.
Jolliffe, I. T. (2002) *Principal Component Analysis, Second Edition*, *Springer*. doi: 10.2307/1270093.
Kalafi, E. Y., Tan, W. B., Town, C. and Dhillon, S. K. (2016) 'Automated identification of Monogeneans using digital image processing and K-nearest neighbour approaches', *BMC Bioinformatics*. BMC Bioinformatics, 17(511), pp. 259–266. doi: 10.1186/s12859-016-1376-z.
Kanan, C. and Cottrell, G. W. (2012) 'Color-to-Grayscale: Does the Method Matter in Image Recognition?', *PLoS ONE*, 7(1), pp. 1–7. doi: 10.1371/.
Kataria, A. and Singh, M. D. (2013) 'A Review of Data Classification Using K-Nearest Neighbour Algorithm', *International Journal of Emerging Technology and Advanced Engineering*, 3(6), pp. 354–360.
Kheysin, Y. M. (2013) *Life Cycles of Coccidia of Domestic Animals*. Edited by J. Kenneth S. Todd. London, UK: William Heinemann Medical Books Limited.
Kohavi, R. (1995) 'A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection', in *The International Joint Conference on Artifcial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 1137–1143. Available at: http://robotics.stanford.edu/~ronnyk.
Kucera, J. and Reznicky, M. (1991) 'Differentiation of species of Eimeria from the fowl using a computerized image-analysis system', *Folia Parasitologica*, 38, pp. 107–113.
Mahajan, S. H. and Harpale, V. K. (2015) 'Adaptive and non-adaptive image interpolation techniques', in *International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015*, pp. 772–775. doi: 10.1109/ICCUBEA.2015.154.
Mehlhorn, H. (2012) *Human Parasites: Diagnosis, Treatment, Prevention*. 7th edn. Dusseldorf, Germany: Spronger. doi: 10.1007/978-3-319-32802-7.
Moallem, P. and Razmjooy, N. (2012) 'Optimal Threshold Computing in Automatic Image Thresholding using Adaptive Particle Swarm Optimization', *Journal of Applied Research and Technology*, 10, pp. 703–712.
N.Suzuki, C. T., Gomes, J. F., Falcao, A. X., Papa, J. P. and Hoshino-Shimizu, S. (2013) 'Automatic segmentation and classification of human intestinal parasites from microscopy images', *IEEE Transactions on Biomedical Engineering*, 60(3), pp. 803–812. doi: 10.1109/TBME.2012.2187204.
Oliveira, U. C., Fraga, J. S., Licois, D., Pakandl, M. and Gruber, A. (2011) 'Development of molecular assays for the identification of the 11 Eimeria species of the domestic rabbit (Oryctolagus cuniculus)', *Veterinary Parasitology*. Elsevier B.V., 176(2–3), pp. 275–280. doi: 10.1016/j.vetpar.2010.10.054.
Otsu, N. (1979) 'A Threshold Selection Method from Gray-Level Histograms', *IEEE transactions on systems*, 9(1), pp. 62–66.
Park, J.-S., Oh, M.-J. and Han, S. (2007) 'Fish Disease Diagnosis System Based on Image Processing of Pathogens' Microscopic Images', in *2007 Frontiers in the Convergence of Bioscience and Information Technologies*, pp. 878–883. doi: 10.1109/FBIT.2007.157.
Peixinho, A. Z., Martins, S. B., Vargas, J. E. and Falcão, A. X. (2016) 'Diagnosis of human intestinal parasites by deep learning', (October 2015), pp. 107–112. doi:

10.1201/b19241-19.
Peng, H., Long, F. and Ding, C. (2005) 'Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy', *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(8), pp. 1226–1238. doi: 10.1109/TPAMI.2005.159.
Plamenov, P. I., Nikolova, K. M. and Georgieva, T. M. (2016) 'On the optimum choice of the K Parameter in Hand-Written Digit Recognition by kNN in comparison to SVM', *INTERNATIONAL JOURNAL OF NEURAL NETWORKS and ADVANCED APPLICATIONS*, 3(Volume 3, 2016), pp. 47–52.
Plitt, A., Imarom, S., Joachim, A. and Daugschies, A. (1999) 'Interactive classification of porcine Eimeria spp. by computer-assisted image analysis', *Veterinary Parasitology*, 86, pp. 105–112. doi: 10.1016/S0304-4017(99)00140-5.
Prajapati, A., Naik, S. and Mehta, S. (2012) 'Evaluation of Different Image Interpolation Algorithms', *International Journal of Computer Applications*, 58(12), pp. 6–12. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Image+Interpolation+and+Resampling#0%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Image+interpolation+and+resampling%230.
Raj, G. D., Aarthi, S., Selvabharathi, R., Raman, M., Blake, D. P. and Tomley, F. M. (2013) 'Real-time PCR-based quantification of Eimeria genomes: A method to outweigh underestimation of genome numbers due to PCR inhibition', *Avian Pathology*, 42(4), pp. 304–308. doi: 10.1080/03079457.2013.790531.
Reddy, P. R., Amarnadh, V. and Bhaskar, M. (2012) 'Evaluation of Stopping Criterion in Contour Tracing Algorithms', *International Journal of Computer Science and Information Technologies*, 3(3), pp. 3888–3894.
Rege, S., Memane, R., Phatak, M. and Agarwal, P. (2013) '2D Geometric Shape and Color Recognition Using Digital Image Processing', *ISSN International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering*, 2(6), pp. 2479–2487. Available at: www.ijareeie.com.
Saif, Y. M. (2008) *Diseases of Poultry*. 12th edn. Edited by A. M. Fadly, J. R. Glisson, L. R. McDougald, L. K. Nolan, and D. E. Swayne. Blackwell Publishing. Available at: www.blackwellprofessional.com.
Saito, P. T. M., Suzuki, C. T. N., Gomes, J. F., De Rezende, P. J. and Falcão, A. X. (2015) 'Robust active learning for the diagnosis of parasites', *Pattern Recognition*. Elsevier, 48(11), pp. 3572–3583. doi: 10.1016/j.patcog.2015.05.020.
Sangeetha, R. and Kalpana, B. (2011) 'Performance Evaluation of Kernels in Multiclass Support Vector Machines', *International Journal of Soft Computing and Engineering (IJSCE)*, 1(5), pp. 138–145. Available at: http://www.ijsce.org/attachments/File/Vol-1_Issue-5/E0171091511.pdf.
Seo, J., Chae, S., Shim, J., Kim, D., Cheong, C. and Han, T. D. (2016) 'Fast contour-tracing algorithm based on a pixel-following method for image sensors', *Sensors (Switzerland)*, 16(3), pp. 1–27. doi: 10.3390/s16030353.
Shah, R. S. (2007) *Support vector machines for classification and regression*, *McGill University Montreal, Quebec*. doi: 10.1039/b918972f.
Sommer, C. (1996) 'Digital image analysis and identification of eggs from bovine

parasitic nematodes', *Journal of Helminthology*, 70(2), p. 143. doi: 10.1017/S0022149X00015303.
Sommer, C. (1998a) 'Quantitative characterization, classification and reconstruction of oocyst shapes of Eimeria species from cattle.', *Parasitology*, 116, pp. 21–28. doi: 10.1017/S003118209700187X.
Sommer, C. (1998b) 'Quantitative characterization of texture used for identification of eggs of bovine parasitic nematodes', *Journal of Helminthology*, 72(2), pp. 179–182. doi: 10.1017/S0022149X00016370.
Stathakis, D. (2009) 'How many hidden layers and nodes?', *International Journal of Remote Sensing*, 30(8), pp. 2133–2147. doi: 10.1080/01431160802549278.
Suzuki, C. T. N., Gomes, J. F., Falcao, A. X., Shimizu, S. H. and Papa, J. P. (2013) 'Automated Diagnosis of Human Intestinal Parasites Using Optical Microscopy Images', in *International Symposium on Biomedical Imaging*. San Francisco, USA, pp. 460–463.
Suzuki, K. (2011) *Artificial Neural Networks Methodological Advances*. 2nd edn. Edited by K. Suzuk. Rijeka, Croatia.
Tchinda, B. S., Tchiotsop, D., Bandjoun, D., Dschang-, U. De and Wolf, D. (2015) 'Automatic Recognition of Human Parasite Cysts on Microscopic Stools Images using Principal Component Analysis and Probabilistic Neural Network', 4(9), pp. 26–33.
Tek_B, F. B., Dempster, A. G. and Kale, I. (2010) 'Parasite detection and identification for automated thin blood film malaria diagnosis', *Computer Vision and Image Understanding*, 114(1), pp. 21–32. doi: 10.1016/j.cviu.2009.08.003.
Tek, F. B. (2007) *Computerised diagnosis of malaria, PhD Thesis*. University of Westminster.
Tek, F. B., Dempster, A. G. and Kale, I. (2009) 'Computer vision for microscopy diagnosis of malaria.', *Malaria journal*, 8(1), p. 153. doi: 10.1186/1475-2875-8-153.
Visa, S., Ramsay, B., Ralescu, A. and Van Der Knaap, E. (2011) 'Confusion matrix-based feature selection', *CEUR Workshop Proceedings*, 710, pp. 120–127.
Wang, L., Wang, Y. and Chang, Q. (2016) 'Feature selection methods for big data bioinformatics: A survey from the search perspective', *Methods*. Elsevier Inc., 111, pp. 21–31. doi: 10.1016/j.ymeth.2016.08.014.
Wiedmer, S., Erdbeer, A., Volke, B., Randel, S., Kapplusch, F., Hanig, S. and Kurth, M. (2017) 'Identification and analysis of Eimeria nieschulzi gametocyte genes reveal splicing events of gam genes and conserved motifs in the wall-forming proteins within the genus Eimeria (Coccidia, Apicomplexa)', *Parasite*, 24(2). doi: 10.1051/parasite/2017049.
Yang, K., Wang, H., Dai, G., Hu, S., Zhang, Y. and Xu, J. (2011) 'Determining the repeat number of cross-validation', in *Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011*, pp. 1706–1710. doi: 10.1109/BMEI.2011.6098566.
Zheng, X., Ye, H. and Tang, Y. (2017) 'Image bi-level thresholding based on gray level-local variance histogram', *Entropy*, 19(5), pp. 1–8. doi: 10.3390/e19050191.
Zonaed Siddiki, A., Mina, S., Anayet Hasan, M., Touaha Akbar, M., Alam, R., Ashraful Islam, M., Zahan Ira, I. and Ayesa, B. (2014) 'Molecular characterization of Eimeria

spp. from chicken by Polymerase Chain Reaction based on species-specific SCAR markers', *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)*, 7(1), pp. 13–17. doi: 10.9790/2380-07111317.

Appendices

Appendix A: Confusion matrices of the best overall classification results for the total features of grey and RGB datasets using K-NN

Table A 1 The confusion matrix of the highest overall accuracy rate of chicken grey CF using K-NN classifier when K=9

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**95.12%**	0.%	0.%	1.38%	3.28%	0.21%	0.01%
	σ	**0.31%**	0.%	0.%	0.21%	0.25%	0.09%	0.03%
BRU	μ	0.%	**84.5%**	0.91%	0.%	0.01%	9.59%	4.98%
	σ	0.%	**0.61%**	0.16%	0.%	0.05%	0.56%	0.33%
MAX	μ	0.%	0.71%	**99.29%**	0.%	0.%	0.%	0.%
	σ	0.%	0.17%	**0.17%**	0.%	0.%	0.%	0.%
MIT	μ	1.29%	0.%	0.%	**92.66%**	2.16%	3.65%	0.24%
	σ	0.17%	0.%	0.%	**0.39%**	0.27%	0.39%	0.11%
NEC	μ	5.59%	0.45%	0.%	6.91%	**58.69%**	11.89%	16.47%
	σ	0.44%	0.1%	0.%	0.5%	**1.07%**	0.79%	0.86%
PRA	μ	0.12%	3.79%	0.%	7.07%	7.14%	**73.43%**	8.44%
	σ	0.05%	0.29%	0.%	0.33%	0.52%	**0.65%**	0.37%
TEN	μ	0.39%	4.58%	0.%	0.43%	14.65%	8.66%	**71.29%**
	σ	0.15%	0.33%	0.%	0.12%	0.69%	0.62%	**0.83%**

Table A 2 The confusion matrix of the highest overall accuracy rate of chicken grey RF using the K-NN classifier when K=15

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**94.57%**	0.%	0.%	1.54%	3.54%	0.32%	0.03%
	σ	**0.23%**	0.%	0.%	0.19%	0.22%	0.11%	0.05%
BRU	μ	0.%	**83.83%**	0.82%	0.%	0.%	10.99%	4.37%
	σ	0.%	**0.58%**	0.11%	0.%	0.%	0.55%	0.35%
MAX	μ	0.%	0.85%	**99.15%**	0.%	0.%	0.%	0.%
	σ	0.%	0.2%	**0.2%**	0.%	0.%	0.%	0.%
MIT	μ	1.29%	0.%	0.%	**92.25%**	1.92%	4.38%	0.16%
	σ	0.2%	0.%	0.%	**0.37%**	0.25%	0.39%	0.1%
NEC	μ	6.35%	0.48%	0.%	9.81%	**43.93%**	17.27%	22.16%
	σ	0.32%	0.06%	0.%	0.44%	**0.97%**	0.83%	0.7%
PRA	μ	0.17%	3.95%	0.%	7.74%	6.01%	**73.34%**	8.8%
	σ	0.06%	0.2%	0.%	0.25%	0.42%	**0.56%**	0.4%
TEN	μ	0.29%	4.95%	0.%	0.54%	14.%	11.1%	**69.11%**
	σ	0.12%	0.31%	0.%	0.15%	0.58%	0.59%	**0.79%**

Table A 3 The confusion matrix of the highest overall accuracy rate of chicken grey CRF using K-NN classifier when K=13

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**96.18%**	0.%	0.%	0.94%	2.87%	0.01%	0.%
	σ	**0.23%**	0.%	0.%	0.09%	0.22%	0.04%	0.%
BRU	μ	0.%	**90.63%**	0.48%	0.%	0.%	4.78%	4.11%
	σ	0.%	**0.67%**	0.07%	0.%	0.%	0.39%	0.44%
MAX	μ	0.%	0.57%	**99.43%**	0.%	0.%	0.%	0.%
	σ	0.%	0.%	**0.%**	0.%	0.%	0.%	0.%
MIT	μ	0.53%	0.%	0.%	**94.77%**	2.1%	2.61%	0.%
	σ	0.13%	0.%	0.%	**0.22%**	0.2%	0.25%	0.%
NEC	μ	10.29%	0.89%	0.%	4.37%	**59.26%**	8.18%	17.%
	σ	0.47%	0.17%	0.%	0.46%	**1.23%**	0.64%	1.1%
PRA	μ	0.56%	4.53%	0.%	6.72%	8.75%	**70.38%**	9.05%
	σ	0.13%	0.27%	0.%	0.22%	0.4%	**0.67%**	0.54%
TEN	μ	0.83%	4.08%	0.%	0.59%	15.7%	6.04%	**72.77%**
	σ	0.16%	0.26%	0.%	0.07%	0.8%	0.66%	**0.86%**

Table A 4 The confusion matrix of the highest overall accuracy rate of chicken RGB CF using the K-NN classifier when K=13

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**97.71%**	0.%	0.%	0.71%	1.58%	0.%	0.%
	σ	**0.21%**	0.%	0.%	0.17%	0.16%	0.%	0.%
BRU	μ	0.%	**97.52%**	0.14%	0.%	0.%	1.17%	1.17%
	σ	0.%	**0.34%**	0.12%	0.%	0.%	0.23%	0.29%
MAX	μ	0.%	0.48%	**99.52%**	0.%	0.%	0.%	0.%
	σ	0.%	0.26%	**0.26%**	0.%	0.%	0.%	0.%
MIT	μ	0.39%	0.%	0.%	**97.53%**	0.88%	1.15%	0.05%
	σ	0.1%	0.%	0.%	**0.28%**	0.17%	0.19%	0.06%
NEC	μ	2.88%	0.%	0.%	4.67%	**73.7%**	9.24%	9.52%
	σ	0.32%	0.%	0.%	0.46%	**0.86%**	0.58%	0.54%
PRA	μ	0.%	0.31%	0.%	1.77%	2.16%	**93.61%**	2.15%
	σ	0.%	0.09%	0.%	0.22%	0.27%	**0.42%**	0.27%
TEN	μ	0.%	0.64%	0.%	0.%	4.92%	2.82%	**91.62%**
	σ	0.%	0.11%	0.%	0.03%	0.4%	0.35%	**0.51%**

Table A 5 The confusion matrix of the highest overall accuracy rate of chicken RGB RF using the K-NN classifier when K=7

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**96.55%**	0.%	0.%	1.11%	1.81%	0.19%	0.34%
	σ	**0.25%**	0.%	0.%	0.14%	0.18%	0.13%	0.1%
BRU	μ	0.%	**97.42%**	0.%	0.%	0.73%	0.1%	1.76%
	σ	0.%	**0.32%**	0.%	0.%	0.12%	0.12%	0.34%
MAX	μ	0.%	0.01%	**99.99%**	0.%	0.%	0.%	0.%
	σ	0.%	0.05%	**0.05%**	0.%	0.%	0.%	0.%
MIT	μ	0.93%	0.%	0.%	**95.91%**	1.95%	1.21%	0.%
	σ	0.11%	0.%	0.%	**0.23%**	0.24%	0.18%	0.%
NEC	μ	5.08%	0.68%	0.%	2.73%	**74.98%**	8.16%	8.36%
	σ	0.56%	0.14%	0.%	0.28%	**1.08%**	0.71%	0.76%
PRA	μ	1.38%	0.64%	0.%	1.83%	7.06%	**71.36%**	17.74%
	σ	0.2%	0.08%	0.%	0.14%	0.43%	**0.79%**	0.65%
TEN	μ	0.8%	0.93%	0.%	0.14%	7.97%	9.05%	**81.12%**
	σ	0.13%	0.07%	0.%	0.01%	0.53%	0.63%	**0.75%**

Table A 6 The confusion matrix of the highest overall accuracy rate of chicken RGB CRF using the K-NN classifier

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**98.71%**	0.%	0.%	0.16%	0.96%	0.05%	0.11%
	σ	**0.3%**	0.%	0.%	0.15%	0.14%	0.11%	0.14%
BRU	μ	0.%	**99.47%**	0.%	0.%	0.%	0.%	0.53%
	σ	0.%	**0.22%**	0.%	0.%	0.%	0.%	0.22%
MAX	μ	0.%	0.06%	**99.94%**	0.%	0.%	0.%	0.%
	σ	0.%	0.12%	**0.12%**	0.%	0.%	0.%	0.%
MIT	μ	0.55%	0.%	0.%	**95.6%**	2.47%	1.38%	0.%
	σ	0.12%	0.%	0.%	**0.43%**	0.29%	0.22%	0.%
NEC	μ	2.7%	0.18%	0.%	1.93%	**82.9%**	6.29%	6.%
	σ	0.52%	0.06%	0.%	0.22%	**0.93%**	0.34%	0.88%
PRA	μ	0.11%	0.12%	0.%	1.79%	3.51%	**91.51%**	2.95%
	σ	0.%	0.04%	0.%	0.15%	0.32%	**0.62%**	0.42%
TEN	μ	0.26%	0.45%	0.%	0.04%	5.23%	3.98%	**90.03%**
	σ	0.15%	0.13%	0.%	0.07%	0.69%	0.62%	**1.11%**

Table A 7 The confusion matrix of the highest overall accuracy rate of rabbit grey CF using K-NN classifier when K=5

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**53.37%**	0.%	5.38%	0.56%	1.19%	4.71%	4.72%	0.%	1.07%	4.46%	24.54%
	σ	**2.17%**	0.%	0.75%	0.2%	0.29%	0.69%	0.91%	0.%	0.58%	0.97%	1.79%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	4.%	0.%	**76.49%**	0.%	0.02%	5.01%	7.01%	0.%	0.44%	3.31%	3.73%
	σ	0.69%	0.%	**1.09%**	0.03%	0.08%	0.56%	0.67%	0.%	0.18%	0.36%	0.73%
INT	μ	1.54%	0.%	0.72%	**66.88%**	0.%	0.%	8.99%	4.64%	13.72%	0.%	3.52%
	σ	0.44%	0.%	0.37%	**1.84%**	0.%	0.%	2.1%	1.06%	1.61%	0.%	0.77%
IRR	μ	4.9%	0.%	1.23%	0.%	**83.26%**	7.02%	0.%	0.%	0.%	3.59%	0.%
	σ	0.78%	0.%	0.51%	0.%	**1.03%**	1.04%	0.%	0.%	0.%	0.74%	0.%
MAG	μ	7.63%	0.%	19.24%	0.%	4.17%	**58.22%**	1.05%	0.%	0.08%	7.38%	2.24%
	σ	0.79%	0.%	0.96%	0.%	0.45%	**1.37%**	0.32%	0.%	0.11%	0.78%	0.5%
MED	μ	4.9%	0.03%	4.62%	1.68%	0.%	0.53%	**74.83%**	2.48%	3.05%	0.32%	7.57%
	σ	1.07%	0.09%	0.85%	0.55%	0.%	0.35%	**1.7%**	0.33%	0.84%	0.31%	1.03%
PER	μ	0.%	0.11%	0.07%	0.25%	0.%	0.%	3.21%	**96.34%**	0.01%	0.%	0.%
	σ	0.%	0.28%	0.31%	0.37%	0.%	0.%	0.93%	**0.96%**	0.07%	0.%	0.%
PIR	μ	3.22%	0.%	0.97%	8.5%	0.%	0.91%	7.46%	0.%	**76.18%**	0.26%	2.5%
	σ	1.12%	0.%	0.54%	1.47%	0.%	0.53%	1.71%	0.%	**1.75%**	0.36%	0.8%
STI	μ	19.6%	0.%	18.81%	0.%	2.01%	13.31%	5.43%	0.%	0.46%	**32.43%**	7.94%
	σ	1.99%	0.%	1.84%	0.%	0.72%	1.83%	1.3%	0.%	0.22%	**2.29%**	1.64%
VEJ	μ	20.65%	0.%	3.63%	1.9%	0.%	0.74%	9.52%	0.%	1.35%	1.02%	**61.19%**
	σ	1.77%	0.%	0.92%	0.47%	0.%	0.39%	1.21%	0.%	0.43%	0.47%	**2.14%**

Table A 8 The confusion matrix of the highest overall accuracy rate of rabbit grey RF using K-NN classifier when K=11

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**41.28%**	1.11%	13.05%	1.72%	0.87%	0.28%	13.02%	0.44%	0.04%	4.4%	23.79%
	σ	**1.08%**	0.46%	0.88%	0.36%	0.%	0.21%	1.18%	0.%	0.13%	0.55%	1.4%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	3.17%	0.%	**73.15%**	1.23%	1.28%	9.69%	0.87%	0.%	1.38%	1.38%	7.88%
	σ	0.46%	0.%	**0.76%**	0.25%	0.41%	0.54%	0.25%	0.%	0.26%	0.3%	0.5%
INT	μ	0.%	0.%	0.82%	**74.58%**	0.%	0.%	3.33%	3.22%	2.61%	0.%	15.43%
	σ	0.%	0.%	0.38%	**1.27%**	0.%	0.%	0.64%	0.85%	0.99%	0.%	0.82%
IRR	μ	0.48%	0.%	11.9%	0.%	**71.51%**	16.07%	0.%	0.%	0.%	0.04%	0.%
	σ	0.18%	0.%	0.97%	0.%	**1.05%**	1.17%	0.%	0.%	0.%	0.15%	0.%
MAG	μ	0.17%	0.%	11.41%	0.%	2.39%	**85.76%**	0.%	0.%	0.%	0.27%	0.%
	σ	0.04%	0.%	0.7%	0.%	0.3%	**0.69%**	0.%	0.%	0.03%	0.18%	0.%
MED	μ	5.86%	0.92%	1.9%	0.89%	0.%	0.%	**65.51%**	3.68%	0.67%	0.03%	20.52%
	σ	0.9%	0.33%	0.39%	0.38%	0.%	0.%	**1.19%**	0.5%	0.2%	0.1%	1.16%
PER	μ	0.%	0.05%	0.%	0.94%	0.%	0.%	7.8%	**91.17%**	0.%	0.%	0.04%
	σ	0.%	0.18%	0.%	0.35%	0.%	0.%	0.5%	**0.47%**	0.%	0.%	0.14%
PIR	μ	0.24%	0.41%	10.63%	12.87%	0.%	0.%	4.%	0.%	**51.19%**	0.%	20.66%
	σ	0.41%	0.37%	1.13%	1.16%	0.%	0.%	0.62%	0.%	**2.21%**	0.%	1.86%
STI	μ	24.19%	0.%	28.66%	0.78%	2.84%	6.68%	0.57%	0.%	0.56%	**18.88%**	16.86%
	σ	1.74%	0.%	1.16%	0.32%	0.53%	0.77%	0.4%	0.%	0.%	**1.28%**	1.18%
VEJ	μ	6.63%	0.%	4.65%	3.9%	0.%	0.%	7.64%	0.%	2.22%	1.12%	**73.85%**
	σ	1.21%	0.%	0.7%	0.52%	0.%	0.%	1.04%	0.%	0.46%	0.45%	**1.69%**

Table A 9 The confusion matrix of the highest overall accuracy rate of rabbit grey CRF using K-NN classifier when K=9

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**57.14%**	0.%	3.46%	0.55%	0.77%	1.04%	4.34%	0.%	0.%	3.3%	29.4%
	σ	**0.97%**	0.%	0.45%	0.%	0.38%	0.41%	0.48%	0.%	0.%	0.58%	1.3%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	1.04%	0.%	**77.65%**	1.74%	0.%	8.9%	1.74%	0.%	1.87%	1.84%	5.21%
	σ	0.32%	0.%	**0.76%**	0.29%	0.%	0.56%	0.14%	0.%	0.31%	0.32%	0.44%
INT	μ	0.%	0.%	0.1%	**90.%**	0.%	0.%	0.%	4.17%	5.%	0.%	0.73%
	σ	0.%	0.%	0.33%	**1.01%**	0.%	0.%	0.%	0.%	1.08%	0.%	0.5%
IRR	μ	1.85%	0.%	0.16%	0.%	**80.27%**	17.39%	0.%	0.%	0.%	0.33%	0.%
	σ	0.38%	0.%	0.26%	0.%	**1.09%**	1.2%	0.%	0.%	0.%	0.28%	0.%
MAG	μ	0.16%	0.%	9.12%	0.%	2.64%	**87.6%**	0.%	0.%	0.%	0.4%	0.07%
	σ	0.14%	0.%	0.36%	0.%	0.25%	**0.54%**	0.%	0.%	0.%	0.14%	0.09%
MED	μ	2.14%	0.%	1.5%	1.23%	0.%	0.%	**79.77%**	1.41%	2.09%	0.86%	11.%
	σ	0.37%	0.%	0.57%	0.43%	0.%	0.%	**1.24%**	0.34%	0.44%	0.14%	1.28%
PER	μ	0.%	0.%	0.%	0.%	0.%	0.%	2.27%	**97.73%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.%	0.%	0.%	0.43%	**0.43%**	0.%	0.%	0.%
PIR	μ	0.%	0.%	2.26%	5.56%	0.%	0.%	4.92%	0.%	**84.92%**	0.%	2.34%
	σ	0.%	0.%	0.64%	0.71%	0.%	0.%	0.6%	0.%	**0.77%**	0.%	0.8%
STI	μ	13.54%	0.%	10.08%	0.79%	0.%	0.63%	0.%	0.%	0.79%	**56.69%**	17.48%
	σ	1.48%	0.%	0.89%	0.%	0.%	0.5%	0.%	0.%	0.%	**1.05%**	0.72%
VEJ	μ	9.8%	0.41%	1.1%	0.73%	0.%	0.%	7.4%	0.%	0.89%	0.81%	**78.86%**
	σ	1.16%	0.%	0.27%	0.17%	0.%	0.%	0.69%	0.%	0.17%	0.%	**1.07%**

Table A 10 The confusion matrix of the highest overall accuracy rate of rabbit RGB CF using K-NN classifier when K=9

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**60.33%**	0.%	4.73%	1.1%	1.98%	0.%	8.74%	0.05%	0.16%	4.73%	18.19%
	σ	**0.68%**	0.%	0.87%	0.26%	0.28%	0.%	0.84%	0.17%	0.27%	0.46%	1.23%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	1.04%	0.%	**71.15%**	1.5%	1.23%	13.66%	1.31%	0.%	2.78%	0.91%	6.42%
	σ	0.27%	0.%	**1.26%**	0.29%	0.34%	0.87%	0.08%	0.%	0.47%	0.4%	0.85%
INT	μ	0.%	0.%	0.%	**91.77%**	0.%	0.%	0.42%	2.08%	2.71%	0.%	3.02%
	σ	0.%	0.%	0.%	**0.59%**	0.%	0.%	0.54%	0.%	0.54%	0.%	0.59%
IRR	μ	0.49%	0.%	4.67%	0.%	**83.42%**	10.38%	0.%	0.%	0.%	1.03%	0.%
	σ	0.17%	0.%	0.38%	0.%	**1.%**	1.04%	0.%	0.%	0.%	0.17%	0.%
MAG	μ	0.%	0.%	10.77%	0.%	4.38%	**84.84%**	0.%	0.%	0.%	0.%	0.02%
	σ	0.%	0.%	0.54%	0.%	0.28%	**0.68%**	0.%	0.%	0.%	0.%	0.06%
MED	μ	6.41%	0.%	0.27%	2.%	0.%	0.%	**76.91%**	1.59%	0.5%	0.45%	11.86%
	σ	0.69%	0.%	0.23%	0.44%	0.%	0.%	**1.19%**	0.24%	0.26%	0.%	0.92%
PER	μ	0.%	0.%	0.%	0.56%	0.%	0.%	3.13%	**96.31%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.16%	0.%	0.%	0.46%	**0.48%**	0.%	0.%	0.%
PIR	μ	0.%	0.%	3.39%	4.44%	0.%	0.%	4.11%	0.%	**75.65%**	0.%	12.42%
	σ	0.%	0.%	1.06%	0.43%	0.%	0.%	0.26%	0.%	**2.18%**	0.%	1.83%
STI	μ	12.44%	0.%	16.46%	0.47%	0.%	0.%	3.46%	0.%	1.1%	**56.85%**	9.21%
	σ	1.22%	0.%	1.2%	0.41%	0.%	0.%	0.66%	0.%	0.41%	**0.97%**	1.44%
VEJ	μ	7.24%	0.41%	1.42%	3.7%	0.%	0.%	5.12%	0.%	0.93%	0.33%	**80.85%**
	σ	1.08%	0.%	0.52%	0.23%	0.%	0.%	0.48%	0.%	0.27%	0.26%	**1.19%**

Table A 11 The confusion matrix of the highest overall accuracy rate of rabbit RGB CRF using K-NN classifier when K=11

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**59.78%**	0.%	3.41%	0.55%	2.97%	2.25%	3.85%	0.%	0.49%	2.2%	24.51%
	σ	**1.32%**	0.%	0.5%	0.%	0.59%	0.31%	0.58%	0.%	0.31%	0.82%	1.42%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	5.45%	0.%	**73.88%**	0.24%	0.05%	4.95%	6.34%	0.%	0.75%	2.99%	5.35%
	σ	0.74%	0.%	**0.83%**	0.15%	0.11%	0.61%	0.52%	0.%	0.21%	0.35%	0.6%
INT	μ	0.%	0.%	0.1%	**83.65%**	0.%	0.%	5.52%	5.21%	4.79%	0.%	0.73%
	σ	0.%	0.%	0.33%	**1.21%**	0.%	0.%	1.3%	0.%	0.54%	0.%	0.7%
IRR	μ	6.9%	0.%	0.22%	0.%	**84.24%**	7.12%	0.%	0.%	0.%	1.3%	0.22%
	σ	1.12%	0.%	0.28%	0.%	**1.09%**	1.48%	0.%	0.%	0.%	0.69%	0.28%
MAG	μ	4.56%	0.%	13.64%	0.%	7.31%	**71.03%**	0.35%	0.%	0.%	2.25%	0.86%
	σ	0.32%	0.%	0.91%	0.%	0.6%	**0.83%**	0.06%	0.%	0.%	0.38%	0.17%
MED	μ	1.82%	0.%	3.86%	0.68%	0.%	0.%	**79.64%**	2.09%	2.5%	1.73%	7.68%
	σ	0.37%	0.%	0.58%	0.69%	0.%	0.%	**1.45%**	0.23%	0.39%	0.47%	0.84%
PER	μ	0.%	0.%	0.%	0.%	0.%	0.%	0.81%	**99.19%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.%	0.%	0.%	0.26%	**0.26%**	0.%	0.%	0.%
PIR	μ	0.%	0.%	0.65%	6.61%	0.%	0.%	4.19%	0.%	**85.24%**	0.16%	3.15%
	σ	0.%	0.%	0.64%	0.34%	0.%	0.%	0.64%	0.%	**0.94%**	0.34%	0.6%
STI	μ	18.5%	0.%	4.88%	0.%	0.%	1.18%	4.88%	0.%	0.63%	**63.15%**	6.77%
	σ	1.75%	0.%	1.1%	0.%	0.%	0.67%	1.22%	0.%	0.33%	**1.1%**	1.3%
VEJ	μ	11.54%	0.41%	2.97%	0.16%	0.%	0.04%	10.45%	0.%	1.26%	0.57%	**72.6%**
	σ	1.27%	0.%	0.67%	0.39%	0.%	0.13%	1.13%	0.%	0.23%	0.21%	**1.5%**

Table A 12 Confusion matrix of the highest overall accuracy of rabbit RGB CRF using the K-NN classifier

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**69.67%**	0.%	1.61%	0.66%	0.15%	0.36%	6.64%	0.%	0.03%	3.5%	17.39%
	σ	**2.%**	0.%	0.61%	0.37%	0.24%	0.17%	1.07%	0.%	0.14%	0.71%	1.91%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	2.34%	0.%	**86.21%**	0.93%	0.64%	2.9%	1.54%	0.%	0.34%	2.8%	2.29%
	σ	0.47%	0.%	**1.04%**	0.24%	0.17%	0.63%	0.35%	0.%	0.16%	0.48%	0.49%
INT	μ	0.52%	0.%	0.%	**79.79%**	0.%	0.%	4.48%	0.93%	9.54%	0.%	4.75%
	σ	0.35%	0.%	0.%	**2.06%**	0.%	0.%	0.9%	0.53%	1.61%	0.%	1.08%
IRR	μ	0.04%	0.%	0.91%	0.%	**89.89%**	7.42%	0.%	0.%	0.%	1.73%	0.%
	σ	0.15%	0.%	0.26%	0.%	**0.94%**	0.87%	0.%	0.%	0.%	0.21%	0.%
MAG	μ	0.15%	0.%	5.67%	0.%	3.58%	**89.67%**	0.%	0.%	0.%	0.92%	0.%
	σ	0.14%	0.%	0.51%	0.%	0.4%	**0.73%**	0.%	0.%	0.%	0.2%	0.%
MED	μ	5.17%	0.%	1.57%	1.23%	0.%	0.%	**80.77%**	1.77%	0.65%	1.75%	7.09%
	σ	0.65%	0.%	0.31%	0.31%	0.%	0.%	**1.09%**	0.55%	0.28%	0.37%	0.66%
PER	μ	0.%	0.%	0.%	0.%	0.%	0.%	1.54%	**98.46%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.%	0.%	0.%	0.2%	**0.2%**	0.%	0.%	0.%
PIR	μ	0.63%	0.%	0.46%	8.13%	0.%	0.%	2.37%	0.%	**85.31%**	0.%	3.1%
	σ	0.36%	0.%	0.61%	1.24%	0.%	0.%	0.83%	0.%	**1.85%**	0.%	0.45%
STI	μ	16.12%	0.%	6.58%	0.58%	0.53%	1.11%	2.62%	0.%	0.%	**66.27%**	6.19%
	σ	2.1%	0.%	1.32%	0.22%	0.64%	0.7%	1.05%	0.%	0.%	**2.84%**	1.17%
VEJ	μ	9.71%	0.%	0.43%	1.65%	0.%	0.%	4.89%	0.%	0.41%	0.84%	**82.08%**
	σ	0.97%	0.%	0.25%	0.51%	0.%	0.%	0.82%	0.%	0.24%	0.33%	**1.4%**

Appendix B: Confusion matrices of the best overall classification results for the total features of grey and RGB datasets using ANN

Table B 1 The confusion matrix of the highest overall accuracy rate of chicken grey CF using ANN

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**94.67%**	0.%	0.%	1.67%	3.25%	0.3%	0.11%
	σ	**0.26%**	0.%	0.%	0.22%	0.21%	0.1%	0.1%
BRU	μ	0.%	**83.11%**	0.89%	0.%	0.%	10.5%	5.49%
	σ	0.%	**1.15%**	0.12%	0.%	0.%	1.28%	0.52%
MAX	μ	0.%	0.59%	**99.38%**	0.%	0.%	0.03%	0.%
	σ	0.%	0.32%	**0.33%**	0.%	0.%	0.08%	0.%
MIT	μ	0.98%	0.%	0.%	**92.25%**	2.68%	3.76%	0.32%
	σ	0.86%	0.08%	0.%	**0.78%**	13.12%	0.97%	4.67%
NEC	μ	3.85%	0.37%	0.%	5.03%	**62.92%**	8.88%	18.95%
	σ	2.%	1.41%	0.%	2.08%	**2.09%**	0.88%	25.26%
PRA	μ	0.24%	3.01%	0.%	5.81%	5.52%	**78.04%**	7.38%
	σ	0.03%	0.43%	0.%	0.53%	1.22%	**2.2%**	1.08%
TEN	μ	0.15%	3.79%	0.%	0.6%	13.06%	8.1%	**74.3%**
	σ	0.17%	0.47%	0.%	0.17%	1.22%	0.98%	**0.%**

Table B 2 The confusion matrix of the highest overall accuracy rate of chicken grey RF using the ANN

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**92.6%**	0.%	0.%	1.68%	5.08%	0.35%	0.29%
	σ	**0.55%**	0.%	0.%	0.12%	0.66%	0.22%	0.17%
BRU	μ	0.%	**85.93%**	0.46%	0.%	0.%	8.81%	4.81%
	σ	0.%	**0.83%**	0.29%	0.%	0.%	0.75%	0.72%
MAX	μ	0.%	0.54%	**99.46%**	0.%	0.%	0.%	0.%
	σ	0.%	0.15%	**0.15%**	0.%	0.%	0.%	0.%
MIT	μ	0.68%	0.%	0.%	**94.43%**	0.79%	4.1%	0.01%
	σ	2.83%	0.3%	0.%	**0.33%**	9.93%	3.99%	4.21%
NEC	μ	10.81%	1.24%	0.%	3.69%	**54.%**	14.44%	15.82%
	σ	3.19%	1.58%	0.%	1.14%	**3.54%**	2.87%	21.45%
PRA	μ	1.03%	4.8%	0.%	6.51%	9.82%	**64.7%**	13.15%
	σ	0.31%	0.73%	0.%	0.35%	1.24%	**2.11%**	1.32%
TEN	μ	1.64%	4.29%	0.%	0.61%	7.98%	19.4%	**66.08%**
	σ	0.25%	0.33%	0.%	0.16%	1.11%	0.73%	**0.03%**

Table B 3 The confusion matrix of the highest overall accuracy rate of chicken grey CRF using ANN

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**95.43%**	0.%	0.%	0.91%	3.44%	0.1%	0.12%
	σ	**0.3%**	0.%	0.%	0.11%	0.23%	0.09%	0.04%
BRU	μ	0.%	**90.48%**	0.3%	0.%	0.%	5.08%	4.14%
	σ	0.%	**0.96%**	0.15%	0.%	0.%	0.55%	0.56%
MAX	μ	0.%	0.54%	**99.43%**	0.%	0.03%	0.%	0.%
	σ	0.%	0.27%	**0.31%**	0.%	0.08%	0.%	0.%
MIT	μ	0.31%	0.%	0.%	**94.87%**	1.61%	3.12%	0.09%
	σ	1.38%	0.1%	0.%	**0.34%**	14.32%	1.48%	4.18%
NEC	μ	5.9%	0.41%	0.%	3.65%	**60.79%**	10.49%	18.76%
	σ	1.86%	1.1%	0.%	1.%	**1.8%**	0.72%	26.95%
PRA	μ	0.39%	2.23%	0.%	5.96%	6.22%	**78.08%**	7.1%
	σ	0.12%	0.3%	0.%	0.4%	0.65%	**1.09%**	0.37%
TEN	μ	0.19%	3.28%	0.%	0.81%	10.32%	6.6%	**78.8%**
	σ	0.11%	0.34%	0.%	0.13%	1.15%	0.5%	**0.%**

Table B 4 The confusion matrix of the highest overall accuracy rate of chicken RGB CF using the ANN

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**97.83%**	0.%	0.%	0.75%	1.43%	0.%	0.%
	σ	**0.48%**	0.%	0.%	0.31%	0.27%	0.%	0.%
BRU	μ	0.%	**98.72%**	0.27%	0.%	0.%	0.43%	0.57%
	σ	0.%	**0.56%**	0.22%	0.%	0.%	0.26%	0.4%
MAX	μ	0.%	0.17%	**99.75%**	0.%	0.%	0.08%	0.%
	σ	0.%	0.23%	**0.24%**	0.%	0.%	0.18%	0.%
MIT	μ	0.59%	0.%	0.%	**96.61%**	1.35%	1.41%	0.04%
	σ	0.46%	0.%	0.%	**0.63%**	17.84%	0.52%	3.23%
NEC	μ	2.03%	0.%	0.%	3.61%	**80.3%**	4.38%	9.68%
	σ	0.56%	0.06%	0.%	1.24%	**2.77%**	0.93%	35.39%
PRA	μ	0.%	0.24%	0.%	1.23%	1.44%	**94.85%**	2.24%
	σ	0.%	0.18%	0.%	0.29%	0.42%	**1.04%**	0.58%
TEN	μ	0.41%	0.4%	0.%	0.01%	5.04%	2.86%	**91.29%**
	σ	0.11%	0.16%	0.%	0.04%	1.21%	0.55%	**0.08%**

Table B 5 The confusion matrix of the highest overall accuracy rate of chicken RGB RF using the ANN

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**92.6%**	0.%	0.%	1.68%	5.08%	0.35%	0.29%
	σ	**0.55%**	0.%	0.%	0.12%	0.66%	0.22%	0.17%
BRU	μ	0.%	**85.93%**	0.46%	0.%	0.%	8.81%	4.81%
	σ	0.%	**0.83%**	0.29%	0.%	0.%	0.75%	0.72%
MAX	μ	0.%	0.54%	**99.46%**	0.%	0.%	0.%	0.%
	σ	0.%	0.15%	**0.15%**	0.%	0.%	0.%	0.%
MIT	μ	0.68%	0.%	0.%	**94.43%**	0.79%	4.1%	0.01%
	σ	2.83%	0.3%	0.%	**0.33%**	9.93%	3.99%	4.21%
NEC	μ	10.81%	1.24%	0.%	3.69%	**54.%**	14.44%	15.82%
	σ	3.19%	1.58%	0.%	1.14%	**3.54%**	2.87%	21.45%
PRA	μ	1.03%	4.8%	0.%	6.51%	9.82%	**64.7%**	13.15%
	σ	0.31%	0.73%	0.%	0.35%	1.24%	**2.11%**	1.32%
TEN	μ	1.64%	4.29%	0.%	0.61%	7.98%	19.4%	**66.08%**
	σ	0.25%	0.33%	0.%	0.16%	1.11%	0.73%	**0.12%**

Table B 6 The confusion matrix of the overall accuracy of chicken RGB CRF using the ANN classifier

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**98.15%**	0.%	0.%	0.38%	1.44%	0.03%	0.%
	σ	**0.33%**	0.%	0.%	0.23%	0.25%	0.05%	0.%
BRU	μ	0.%	**99.22%**	0.07%	0.%	0.02%	0.07%	0.62%
	σ	0.%	**0.27%**	0.21%	0.%	0.07%	0.1%	0.21%
MAX	μ	0.%	0.25%	**99.75%**	0.%	0.%	0.%	0.%
	σ	0.%	0.35%	**0.35%**	0.%	0.%	0.%	0.%
MIT	μ	0.23%	0.%	0.%	**97.74%**	0.93%	1.09%	0.%
	σ	0.36%	0.%	0.%	**0.18%**	20.09%	0.89%	1.89%
NEC	μ	2.03%	0.%	0.%	1.83%	**84.85%**	3.45%	7.85%
	σ	0.74%	0.18%	0.%	0.6%	**3.05%**	0.34%	36.1%
PRA	μ	0.%	0.06%	0.%	1.26%	1.57%	**95.45%**	1.67%
	σ	0.%	0.08%	0.%	0.47%	0.44%	**1.36%**	0.53%
TEN	μ	0.4%	0.29%	0.%	0.06%	3.63%	1.8%	**93.82%**
	σ	0.34%	0.11%	0.%	0.07%	1.2%	0.57%	**0.01%**

Table B 7 The confusion matrix of the highest overall accuracy rate of rabbit grey CF using ANN

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**44.45%**	0.17%	2.93%	0.74%	1.%	6.77%	5.5%	0.%	1.62%	4.32%	32.49%
	σ	**3.97%**	0.52%	0.87%	0.98%	0.52%	1.74%	1.13%	0.%	0.52%	1.66%	2.92%
EXI	μ	0.%	**99.97%**	0.%	0.%	0.%	0.%	0.%	0.03%	0.%	0.%	0.%
	σ	0.%	**0.1%**	0.%	0.%	0.%	0.%	0.%	0.1%	0.%	0.%	0.%
FLA	μ	0.94%	0.%	**73.32%**	0.19%	0.24%	13.03%	4.99%	0.%	0.31%	3.03%	3.95%
	σ	0.57%	0.%	**2.74%**	0.1%	0.24%	2.26%	0.76%	0.%	0.11%	0.74%	0.58%
INT	μ	0.22%	0.%	0.07%	**80.37%**	0.%	0.%	3.21%	3.28%	9.63%	0.%	3.21%
	σ	0.48%	0.%	0.22%	**1.77%**	0.%	0.%	1.%	1.3%	1.08%	0.%	1.%
IRR	μ	1.47%	0.%	0.27%	0.%	**88.21%**	8.37%	0.%	0.%	0.%	1.68%	0.%
	σ	1.09%	0.%	0.27%	0.%	**1.95%**	1.17%	0.%	0.%	0.%	0.62%	0.%
MAG	μ	1.44%	0.%	11.17%	0.%	2.59%	**78.94%**	0.05%	0.%	0.%	4.55%	1.26%
	σ	0.35%	0.%	1.41%	0.%	0.32%	**1.82%**	0.12%	0.%	0.%	1.%	0.16%
MED	μ	1.71%	0.1%	5.55%	0.75%	0.03%	0.45%	**80.27%**	2.4%	0.89%	1.88%	5.96%
	σ	0.72%	0.16%	1.44%	0.26%	0.1%	0.53%	**2.01%**	0.85%	0.35%	0.56%	1.02%
PER	μ	0.%	0.05%	0.%	0.15%	0.%	0.%	3.18%	**96.62%**	0.%	0.%	0.%
	σ	0.%	0.15%	0.%	0.23%	0.%	0.%	1.08%	**1.01%**	0.%	0.%	0.%
PIR	μ	0.44%	0.%	0.88%	7.79%	0.%	0.37%	2.43%	0.%	**85.59%**	0.22%	2.28%
	σ	0.36%	0.%	0.55%	1.92%	0.%	0.37%	0.34%	0.%	**1.65%**	0.34%	0.69%
STI	μ	5.5%	0.%	12.5%	0.33%	1.%	28.72%	1.33%	0.%	0.44%	**43.11%**	7.06%
	σ	2.1%	0.%	2.62%	0.27%	0.6%	3.51%	0.37%	0.%	0.48%	**5.11%**	1.45%
VEJ	μ	10.21%	0.%	2.36%	0.76%	0.%	0.76%	8.06%	0.%	1.22%	1.18%	**75.45%**
	σ	3.02%	0.%	0.93%	0.3%	0.%	0.4%	2.09%	0.%	0.59%	0.44%	**4.16%**

Table B 8 The confusion matrix of the highest overall accuracy rate of rabbit grey RF using ANN

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**38.12%**	0.26%	10.35%	2.23%	0.87%	0.22%	14.76%	0.44%	0.96%	13.06%	18.73%
	σ	**4.14%**	0.56%	1.12%	0.66%	0.%	0.29%	2.02%	0.%	0.61%	2.8%	1.74%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	1.69%	0.%	**65.67%**	1.02%	2.25%	16.78%	0.41%	0.%	2.3%	2.47%	7.41%
	σ	0.54%	0.%	**2.99%**	0.28%	0.48%	2.74%	0.29%	0.%	0.73%	0.9%	0.88%
INT	μ	0.%	0.%	0.45%	**76.12%**	0.%	0.%	3.28%	2.24%	6.27%	0.%	11.64%
	σ	0.%	0.%	0.37%	**1.45%**	0.%	0.%	1.01%	0.58%	1.17%	0.%	1.57%
IRR	μ	0.71%	0.%	4.13%	0.%	**79.95%**	15.11%	0.%	0.%	0.%	0.11%	0.%
	σ	0.49%	0.%	1.36%	0.%	**1.69%**	1.66%	0.%	0.%	0.%	0.22%	0.%
MAG	μ	0.04%	0.%	7.%	0.%	3.65%	**88.94%**	0.%	0.%	0.18%	0.2%	0.%
	σ	0.07%	0.%	1.3%	0.%	0.48%	**1.53%**	0.%	0.%	0.%	0.15%	0.%
MED	μ	3.18%	0.24%	1.47%	1.3%	0.%	0.%	**75.%**	3.63%	0.82%	1.13%	13.22%
	σ	1.4%	0.41%	0.31%	0.21%	0.%	0.%	**1.64%**	0.41%	0.31%	0.43%	1.85%
PER	μ	0.%	0.%	0.%	0.56%	0.%	0.%	8.64%	**90.71%**	0.%	0.%	0.1%
	σ	0.%	0.%	0.%	0.53%	0.%	0.%	1.57%	**1.55%**	0.%	0.%	0.2%
PIR	μ	0.22%	0.07%	8.31%	8.31%	0.%	0.%	1.84%	0.59%	**73.75%**	0.%	6.91%
	σ	0.47%	0.22%	2.44%	1.4%	0.%	0.%	0.49%	0.29%	**2.11%**	0.%	1.92%
STI	μ	16.56%	0.%	23.44%	0.83%	3.67%	6.17%	0.78%	0.%	0.44%	**36.78%**	11.33%
	σ	3.09%	0.%	2.6%	0.37%	1.52%	1.04%	0.44%	0.%	0.22%	**3.85%**	1.06%
VEJ	μ	5.38%	0.%	7.08%	4.79%	0.%	0.%	11.18%	0.%	1.84%	3.02%	**66.7%**
	σ	1.26%	0.%	1.45%	0.8%	0.%	0.%	1.46%	0.%	0.78%	0.81%	**2.2%**

Table B 9 The confusion matrix of the highest overall accuracy rate of rabbit grey CRF using ANN

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**57.16%**	0.%	5.81%	0.17%	0.74%	0.44%	5.85%	0.%	0.96%	5.28%	23.58%
	σ	**5.51%**	0.%	1.6%	0.35%	0.65%	0.44%	0.86%	0.%	0.61%	1.94%	3.16%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.99%	0.%	**82.93%**	0.48%	0.24%	7.34%	2.57%	0.%	0.77%	3.63%	1.04%
	σ	0.37%	0.%	**2.54%**	0.24%	0.24%	1.42%	0.39%	0.%	0.34%	1.19%	0.48%
INT	μ	0.%	0.%	0.22%	**85.52%**	0.%	0.%	1.72%	1.19%	8.21%	0.%	3.13%
	σ	0.%	0.%	0.34%	**2.34%**	0.%	0.%	1.21%	1.54%	1.45%	0.%	0.56%
IRR	μ	0.%	0.%	0.05%	0.%	**90.6%**	9.13%	0.%	0.%	0.%	0.22%	0.%
	σ	0.%	0.%	0.16%	0.%	**1.06%**	1.13%	0.%	0.%	0.%	0.27%	0.%
MAG	μ	0.%	0.%	4.82%	0.%	1.89%	**92.61%**	0.%	0.%	0.%	0.68%	0.%
	σ	0.%	0.%	0.79%	0.%	0.23%	**0.9%**	0.%	0.%	0.%	0.26%	0.%
MED	μ	2.09%	0.07%	3.18%	0.68%	0.%	0.%	**85.41%**	1.99%	0.45%	1.58%	4.55%
	σ	0.9%	0.14%	0.51%	0.31%	0.%	0.%	**1.64%**	0.55%	0.27%	0.46%	1.08%
PER	μ	0.%	0.%	0.%	0.%	0.%	0.%	3.03%	**96.97%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.%	0.%	0.%	1.08%	**1.08%**	0.%	0.%	0.%
PIR	μ	0.07%	0.%	2.57%	8.9%	0.%	0.%	2.13%	0.15%	**83.75%**	0.29%	2.13%
	σ	0.22%	0.%	0.75%	2.17%	0.%	0.%	0.4%	0.29%	**2.04%**	0.36%	0.61%
STI	μ	10.17%	0.%	14.39%	0.44%	2.22%	5.39%	0.56%	0.%	0.33%	**60.22%**	6.28%
	σ	3.68%	0.%	1.54%	0.22%	0.66%	1.22%	0.61%	0.%	0.37%	**4.85%**	1.06%
VEJ	μ	10.1%	0.%	1.53%	1.39%	0.%	0.%	5.97%	0.%	0.28%	1.67%	**79.06%**
	σ	2.56%	0.%	0.54%	0.6%	0.%	0.%	1.43%	0.%	0.34%	0.4%	**2.65%**

Table B 10 The confusion matrix of the highest overall accuracy rate of rabbit RGB CF using ANN

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**77.51%**	0.%	2.36%	0.48%	0.79%	0.61%	4.67%	0.%	0.09%	3.23%	10.26%
	σ	**4.93%**	0.%	0.88%	0.6%	0.51%	0.59%	3.19%	0.%	0.26%	1.%	1.97%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.75%	0.%	**89.78%**	0.29%	0.12%	3.1%	2.18%	0.%	0.27%	1.65%	1.86%
	σ	0.44%	0.%	**1.82%**	0.15%	0.29%	0.76%	1.16%	0.%	0.3%	0.84%	0.5%
INT	μ	0.15%	0.07%	0.%	**95.%**	0.%	0.%	0.82%	0.67%	2.39%	0.15%	0.75%
	σ	0.3%	0.22%	0.%	**2.31%**	0.%	0.%	0.97%	0.62%	2.05%	0.3%	0.82%
IRR	μ	0.43%	0.%	0.22%	0.%	**96.25%**	2.5%	0.%	0.%	0.%	0.22%	0.38%
	σ	0.33%	0.%	0.36%	0.%	**1.13%**	0.65%	0.%	0.%	0.%	0.27%	0.25%
MAG	μ	0.13%	0.%	2.37%	0.04%	1.04%	**94.96%**	0.13%	0.%	0.%	0.77%	0.56%
	σ	0.14%	0.%	0.65%	0.07%	0.38%	**1.31%**	0.12%	0.%	0.%	0.5%	0.56%
MED	μ	2.6%	0.07%	2.26%	0.21%	0.%	0.14%	**89.49%**	1.99%	0.31%	0.55%	2.4%
	σ	1.11%	0.14%	1.28%	0.41%	0.%	0.23%	**2.26%**	0.59%	0.32%	0.53%	1.22%
PER	μ	0.%	0.%	0.%	0.05%	0.%	0.%	1.87%	**98.08%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.15%	0.%	0.%	1.28%	**1.35%**	0.%	0.%	0.%
PIR	μ	0.37%	0.%	0.44%	4.49%	0.%	0.%	0.51%	0.07%	**92.5%**	0.29%	1.32%
	σ	0.37%	0.%	0.94%	2.14%	0.%	0.%	0.47%	0.22%	**2.18%**	0.36%	0.55%
STI	μ	5.39%	0.%	5.33%	0.11%	0.22%	3.22%	1.11%	0.%	0.06%	**83.78%**	0.78%
	σ	2.24%	0.%	2.05%	0.22%	0.44%	1.7%	1.47%	0.%	0.17%	**4.17%**	0.67%
VEJ	μ	6.39%	0.%	1.22%	0.8%	0.%	0.1%	3.78%	0.%	0.14%	0.17%	**87.4%**
	σ	1.2%	0.%	0.64%	0.52%	0.%	0.22%	1.06%	0.%	0.23%	0.32%	**1.72%**

Table B 11 The confusion matrix of the highest overall accuracy rate of rabbit RGB CRF using ANN

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**65.76%**	0.09%	3.06%	2.1%	1.14%	0.17%	10.66%	0.%	0.26%	4.8%	11.97%
	σ	**5.65%**	0.26%	0.85%	1.45%	0.45%	0.35%	3.38%	0.%	0.35%	1.44%	3.14%
EXI	μ	0.%	**99.97%**	0.%	0.03%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.1%**	0.%	0.1%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.27%	0.%	**90.58%**	0.34%	0.31%	4.24%	0.77%	0.02%	1.14%	0.87%	1.45%
	σ	0.23%	0.%	**2.54%**	0.22%	0.27%	1.5%	1.15%	0.07%	0.68%	0.6%	0.63%
INT	μ	0.82%	0.22%	0.%	**89.55%**	0.%	0.%	3.88%	0.75%	2.09%	0.97%	1.72%
	σ	0.78%	0.34%	0.%	**2.14%**	0.%	0.%	1.94%	0.75%	1.85%	0.89%	0.82%
IRR	μ	0.43%	0.%	1.25%	0.%	**94.95%**	2.55%	0.%	0.%	0.%	0.54%	0.27%
	σ	0.41%	0.%	1.%	0.%	**1.24%**	1.%	0.%	0.%	0.%	0.54%	0.27%
MAG	μ	0.04%	0.%	3.42%	0.%	1.4%	**94.55%**	0.02%	0.%	0.%	0.4%	0.18%
	σ	0.07%	0.%	0.75%	0.%	0.38%	**1.29%**	0.05%	0.%	0.%	0.49%	0.43%
MED	μ	4.21%	0.07%	1.06%	1.13%	0.%	0.07%	**86.2%**	2.12%	0.17%	1.95%	3.01%
	σ	1.41%	0.14%	1.49%	0.74%	0.%	0.14%	**2.49%**	0.76%	0.23%	1.05%	0.59%
PER	μ	0.%	0.%	0.%	0.05%	0.%	0.%	3.13%	**96.82%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.15%	0.%	0.%	1.49%	**1.48%**	0.%	0.%	0.%
PIR	μ	0.22%	0.07%	2.57%	2.28%	0.%	0.%	0.51%	0.07%	**92.57%**	0.07%	1.62%
	σ	0.34%	0.22%	2.01%	2.78%	0.%	0.%	0.57%	0.22%	**4.43%**	0.22%	1.03%
STI	μ	4.72%	0.%	3.56%	0.44%	0.56%	1.78%	2.94%	0.%	0.06%	**85.61%**	0.33%
	σ	1.61%	0.%	2.24%	0.22%	0.66%	1.77%	1.27%	0.%	0.17%	**4.58%**	0.57%
VEJ	μ	4.79%	0.%	0.97%	1.84%	0.03%	0.%	4.24%	0.%	0.52%	0.%	**87.6%**
	σ	1.54%	0.%	0.99%	1.22%	0.1%	0.%	1.66%	0.%	0.32%	0.%	**2.31%**

Table B 12 Confusion matrix of the overall accuracy of rabbit RGB CRF using the ANN classifier

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**82.75%**	0.%	1.44%	0.17%	0.7%	0.13%	4.89%	0.%	0.26%	1.83%	7.82%
	σ	**4.46%**	0.%	0.71%	0.29%	0.49%	0.28%	1.73%	0.%	0.4%	0.78%	2.77%
EXI	μ	0.%	**99.97%**	0.%	0.03%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.1%**	0.%	0.1%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.17%	0.%	**95.71%**	0.24%	0.02%	1.74%	0.44%	0.%	0.17%	0.7%	0.8%
	σ	0.24%	0.%	**1.4%**	0.11%	0.07%	0.79%	0.18%	0.%	0.16%	0.45%	0.31%
INT	μ	0.07%	0.%	0.%	**94.33%**	0.%	0.%	2.39%	0.07%	1.49%	0.%	1.64%
	σ	0.22%	0.%	0.%	**2.59%**	0.%	0.%	1.15%	0.22%	1.2%	0.%	1.33%
IRR	μ	0.16%	0.%	0.%	0.%	**96.96%**	2.88%	0.%	0.%	0.%	0.%	0.%
	σ	0.25%	0.%	0.%	0.%	**1.48%**	1.35%	0.%	0.%	0.%	0.%	0.%
MAG	μ	0.%	0.%	1.46%	0.02%	0.7%	**97.73%**	0.%	0.%	0.%	0.07%	0.02%
	σ	0.%	0.%	0.55%	0.05%	0.31%	**0.63%**	0.%	0.%	0.%	0.12%	0.05%
MED	μ	2.02%	0.%	0.34%	0.41%	0.%	0.%	**92.67%**	1.75%	0.1%	0.51%	2.19%
	σ	0.68%	0.%	0.27%	0.21%	0.%	0.%	**0.83%**	0.66%	0.16%	0.28%	0.71%
PER	μ	0.%	0.05%	0.%	0.%	0.%	0.%	1.36%	**98.59%**	0.%	0.%	0.%
	σ	0.%	0.15%	0.%	0.%	0.%	0.%	1.17%	**1.12%**	0.%	0.%	0.%
PIR	μ	0.%	0.%	0.59%	1.76%	0.%	0.07%	0.29%	0.%	**95.37%**	0.%	1.91%
	σ	0.%	0.%	0.44%	1.%	0.%	0.22%	0.36%	0.%	**1.68%**	0.%	0.75%
STI	μ	3.22%	0.%	2.06%	0.%	0.22%	0.89%	0.56%	0.%	0.%	**92.83%**	0.22%
	σ	2.39%	0.%	1.14%	0.%	0.44%	0.94%	0.61%	0.%	0.%	**3.6%**	0.27%
VEJ	μ	5.97%	0.%	0.73%	0.8%	0.%	0.%	2.78%	0.%	0.1%	0.17%	**89.44%**
	σ	0.86%	0.%	0.33%	0.38%	0.%	0.%	0.69%	0.%	0.22%	0.23%	**1.77%**

Appendix C: Confusion matrices of the best overall classification results for the total features of grey and RGB datasets using SVM

Table C 1 The confusion matrix of the highest overall accuracy rate of chicken grey CF using SVM

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**94.42%**	0.%	0.%	1.07%	3.79%	0.14%	0.07%
	σ	**0.%**	0.%	0.%	0.04%	0.17%	0.%	0.07%
BRU	μ	0.%	**84.99%**	0.48%	0.%	0.46%	4.58%	5.9%
	σ	0.%	**0.25%**	0.16%	0.%	0.%	0.31%	0.25%
MAX	μ	0.%	1.39%	**99.41%**	0.28%	0.28%	0.%	0.%
	σ	0.%	0.08%	**0.2%**	0.%	0.%	0.%	0.%
MIT	μ	1.35%	0.%	0.%	**92.48%**	4.49%	7.7%	0.53%
	σ	0.84%	0.%	0.%	**0.15%**	11.95%	0.54%	3.73%
NEC	μ	5.66%	0.%	0.%	4.36%	**55.84%**	10.69%	16.75%
	σ	1.7%	1.1%	0.%	1.27%	**0.88%**	1.71%	27.15%
PRA	μ	0.25%	4.77%	0.%	3.35%	5.17%	**75.99%**	5.45%
	σ	0.05%	0.18%	0.%	0.1%	0.3%	**0.3%**	0.28%
TEN	μ	0.%	2.69%	0.%	0.15%	15.29%	11.23%	**76.35%**
		0.%	0.15%	0.%	0.%	0.34%	0.33%	**0.76%**

Table C 2 The confusion matrix of the highest overall accuracy rate of chicken grey RF using the SVM

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**94.12%**	0.%	0.%	1.22%	9.1%	1.77%	1.45%
	σ	**0.09%**	0.%	0.%	0.14%	0.14%	0.09%	0.11%
BRU	μ	0.%	**88.74%**	0.23%	0.%	1.65%	7.73%	8.03%
	σ	0.%	**0.81%**	0.%	0.%	0.14%	0.41%	0.4%
MAX	μ	0.%	0.74%	**99.72%**	0.28%	0.%	0.%	0.%
	σ	0.%	0.14%	**0.%**	0.%	0.%	0.%	0.%
MIT	μ	1.45%	0.%	0.%	**94.%**	2.29%	6.49%	0.49%
	σ	0.65%	0.%	0.%	**0.24%**	10.79%	0.71%	3.45%
NEC	μ	5.15%	0.%	0.%	1.91%	**47.93%**	13.49%	14.65%
	σ	1.5%	1.34%	0.%	0.61%	**0.44%**	1.43%	25.98%
PRA	μ	0.56%	2.71%	0.%	3.31%	9.3%	**71.57%**	7.44%
	σ	0.%	0.28%	0.%	0.15%	0.3%	**0.49%**	0.16%
TEN	μ	0.16%	3.31%	0.%	0.%	11.92%	12.61%	**72.47%**
	σ	0.04%	0.29%	0.%	0.%	0.29%	0.48%	**0.63%**

Table C 3 The confusion matrix of the highest overall accuracy rate of chicken grey CRF using SVM

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**95.91%**	0.%	0.%	0.53%	4.46%	0.41%	0.16%
	σ	**0.25%**	0.%	0.%	0.11%	0.32%	0.09%	0.05%
BRU	μ	0.%	**90.41%**	0.05%	0.%	0.46%	4.39%	5.4%
	σ	0.%	**0.59%**	0.09%	0.%	0.%	0.25%	0.21%
MAX	μ	0.%	2.18%	**99.94%**	0.28%	0.57%	0.4%	0.34%
	σ	0.%	0.42%	**0.11%**	0.%	0.%	0.14%	0.17%
MIT	μ	0.82%	0.%	0.%	**94.24%**	1.7%	6.57%	0.46%
	σ	0.77%	0.%	0.%	**0.13%**	15.33%	0.2%	3.41%
NEC	μ	4.46%	0.%	0.%	3.02%	**65.66%**	10.14%	15.11%
	σ	1.32%	0.91%	0.%	0.92%	**0.85%**	0.47%	29.1%
PRA	μ	0.07%	2.15%	0.%	3.04%	4.5%	**79.06%**	3.44%
	σ	0.07%	0.17%	0.%	0.11%	0.24%	**0.2%**	0.26%
TEN	μ	0.12%	2.21%	0.%	0.%	11.54%	8.59%	**80.28%**
	σ	0.06%	0.1%	0.%	0.%	0.43%	0.34%	**0.71%**

Table C 4 The confusion matrix of the highest overall accuracy rate of chicken RGB CF using the SVM

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**97.98%**	0.%	0.%	0.34%	1.15%	0.%	0.%
	σ	**0.22%**	0.%	0.%	0.09%	0.07%	0.%	0.%
BRU	μ	0.%	**98.56%**	0.%	0.%	0.%	0.62%	0.85%
	σ	0.%	**0.23%**	0.%	0.%	0.%	0.1%	0.1%
MAX	μ	0.62%	1.56%	**100.%**	0.28%	0.62%	1.25%	0.71%
	σ	0.17%	0.19%	**0.%**	0.%	0.17%	0.14%	0.14%
MIT	μ	0.54%	0.%	0.%	**97.36%**	2.87%	1.35%	0.%
	σ	0.29%	0.%	0.%	**0.17%**	18.29%	0.26%	1.%
NEC	μ	1.68%	0.%	0.%	1.56%	**80.41%**	2.52%	3.77%
	σ	0.52%	0.12%	0.%	0.44%	**0.61%**	0.23%	38.35%
PRA	μ	0.%	0.01%	0.%	1.09%	2.39%	**95.11%**	1.26%
	σ	0.%	0.03%	0.%	0.09%	0.18%	**0.42%**	0.12%
TEN	μ	0.%	0.1%	0.%	0.09%	6.06%	1.89%	**94.73%**
	σ	0.%	0.09%	0.%	0.07%	0.37%	0.18%	**0.52%**

Table C 5 The confusion matrix of the highest overall accuracy rate of chicken RGB RF using the SVM

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**96.51%**	0.%	0.%	0.56%	1.55%	0.58%	0.39%
	σ	**0.24%**	0.%	0.%	0.1%	0.27%	0.06%	0.15%
BRU	μ	0.%	**93.07%**	0.%	0.%	1.12%	0.%	1.56%
	σ	0.%	**0.45%**	0.%	0.%	0.07%	0.%	0.14%
MAX	μ	1.08%	7.03%	**100.%**	0.74%	1.61%	2.21%	1.19%
	σ	0.11%	0.44%	**0.%**	0.14%	0.46%	0.11%	0.21%
MIT	μ	1.13%	0.%	0.%	**96.73%**	1.43%	1.76%	0.%
	σ	0.18%	0.08%	0.%	**0.14%**	18.88%	0.42%	1.01%
NEC	μ	2.27%	0.22%	0.%	1.87%	**79.61%**	4.5%	4.3%
	σ	0.69%	0.12%	0.%	0.59%	**1.08%**	1.72%	34.73%
PRA	μ	0.17%	0.23%	0.%	1.21%	4.25%	**89.21%**	6.65%
	σ	0.1%	0.%	0.%	0.05%	0.31%	**0.4%**	0.29%
TEN	μ	0.%	0.35%	0.%	0.%	4.28%	6.9%	**86.22%**
	σ	0.%	0.1%	0.%	0.%	0.4%	0.27%	**0.59%**

Table C 6 The confusion matrix of the overall accuracy for chicken RGB CRF using SVM classifier

		ACE	BRU	MAX	MIT	NEC	PRA	TEN
ACE	μ	**98.48%**	0.%	0.%	0.08%	1.33%	0.%	0.%
	σ	**0.34%**	0.%	0.%	0.11%	0.23%	0.%	0.%
BRU	μ	0.%	**98.86%**	0.%	0.%	0.%	0.23%	1.28%
	σ	0.%	**0.52%**	0.%	0.%	0.%	0.2%	0.37%
MAX	μ	0.%	0.06%	**99.26%**	0.%	0.%	0.%	0.%
	σ	0.%	0.11%	**0.61%**	0.%	0.%	0.%	0.%
MIT	μ	0.42%	0.%	0.%	**97.64%**	1.03%	1.87%	0.02%
	σ	0.16%	0.%	0.%	**0.14%**	24.91%	0.39%	0.99%
NEC	μ	1.1%	0.%	0.%	1.62%	**85.52%**	4.22%	4.02%
	σ	0.52%	0.24%	0.%	0.66%	**0.66%**	0.81%	49.91%
PRA	μ	0.27%	0.27%	0.29%	1.2%	2.44%	**94.52%**	2.03%
	σ	0.09%	0.23%	0.24%	0.06%	0.21%	**0.43%**	0.6%
TEN	μ	0.%	0.35%	0.%	0.%	4.63%	1.7%	**93.62%**
	σ	0.%	0.07%	0.%	0.%	0.22%	0.18%	**0.%**

Table C 7 The confusion matrix of the highest overall accuracy rate of rabbit grey CF using SVM

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**51.4%**	0.%	2.18%	0.22%	0.66%	4.02%	2.01%	0.%	0.%	6.2%	15.81%
	σ	**1.59%**	0.%	0.2%	0.22%	0.29%	0.47%	0.21%	0.%	0.%	0.61%	0.8%
EXI	μ	0.%	**99.66%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	1.6%	0.%	**77.34%**	0.%	0.%	10.39%	3.54%	0.%	0.24%	5.96%	1.26%
	σ	0.39%	0.%	**0.55%**	0.%	0.%	0.64%	0.33%	0.%	0.%	0.19%	0.18%
INT	μ	0.6%	0.%	0.%	**76.87%**	0.%	0.%	2.76%	0.15%	8.73%	0.%	0.75%
	σ	0.3%	0.%	0.%	**2.19%**	0.%	0.%	0.34%	0.45%	1.16%	0.%	0.%
IRR	μ	1.47%	0.%	0.%	0.%	**82.61%**	5.49%	0.%	0.%	0.%	1.14%	0.%
	σ	0.25%	0.%	0.%	0.%	**1.29%**	0.29%	0.%	0.%	0.%	0.38%	0.%
MAG	μ	3.72%	0.18%	7.3%	0.05%	5.41%	**84.64%**	0.32%	0.%	0.04%	7.23%	0.49%
	σ	0.33%	0.%	0.44%	0.12%	0.54%	**0.78%**	0.11%	0.%	0.07%	0.55%	0.14%
MED	μ	2.09%	0.%	7.16%	1.64%	0.%	0.%	**81.3%**	2.02%	1.47%	0.45%	8.08%
	σ	0.5%	0.%	0.45%	0.63%	0.%	0.%	**0.91%**	0.1%	0.22%	0.16%	0.51%
PER	μ	0.%	0.%	0.%	2.07%	0.%	0.%	2.88%	**96.92%**	0.15%	0.%	0.%
	σ	0.%	0.%	0.%	0.15%	0.%	0.%	0.32%	**0.35%**	0.23%	0.%	0.%
PIR	μ	2.43%	0.%	0.%	12.72%	0.%	0.%	1.47%	0.%	**84.56%**	0.74%	0.29%
	σ	0.34%	0.%	0.%	0.99%	0.%	0.%	0.57%	0.%	**0.66%**	0.%	0.49%
STI	μ	2.61%	0.%	4.83%	0.%	0.22%	9.22%	0.72%	0.%	0.56%	**47.17%**	0.61%
	σ	0.5%	0.%	1.06%	0.%	0.27%	1.06%	0.25%	0.%	0.%	**1.39%**	0.17%
VEJ	μ	23.06%	0.%	6.39%	1.39%	0.%	2.29%	7.26%	0.%	0.87%	4.06%	**75.63%**
	σ	0.54%	0.%	0.28%	0.%	0.%	0.42%	0.81%	0.%	0.17%	0.38%	**0.74%**

Table C 8 The confusion matrix of the highest overall accuracy rate of rabbit grey RF using SVM

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**42.05%**	0.%	4.1%	0.17%	0.17%	0.35%	7.12%	0.09%	0.%	9.96%	9.3%
	σ	**1.11%**	0.%	2.91%	0.35%	0.35%	0.17%	5.01%	0.17%	0.%	0.95%	5.67%
EXI	μ	0.%	**99.66%**	0.%	0.07%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.14%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	4.43%	0.%	**71.45%**	0.53%	1.79%	9.93%	0.99%	0.%	2.54%	8.52%	4.55%
	σ	1.69%	0.%	**0.81%**	0.36%	0.46%	1.69%	0.33%	0.%	0.83%	3.37%	1.69%
INT	μ	1.34%	0.52%	3.36%	**75.37%**	0.%	0.%	1.87%	1.12%	8.96%	0.6%	8.43%
	σ	0.73%	0.34%	0.76%	**0.94%**	0.%	0.%	0.6%	0.6%	2.38%	0.3%	2.81%
IRR	μ	0.87%	0.%	2.88%	0.%	**75.38%**	9.02%	0.%	0.%	0.%	2.88%	0.%
	σ	0.43%	0.%	0.97%	0.%	**0.65%**	5.7%	0.%	0.%	0.%	1.46%	0.%
MAG	μ	0.13%	0.%	9.59%	0.%	5.76%	**90.86%**	0.14%	0.02%	0.14%	1.35%	0.%
	σ	0.08%	0.%	1.56%	0.%	1.82%	**0.42%**	0.07%	0.05%	0.07%	0.68%	0.%
MED	μ	10.92%	0.%	0.72%	1.3%	0.%	0.07%	**76.1%**	4.66%	1.68%	0.48%	10.1%
	σ	3.52%	0.%	0.47%	0.34%	0.%	0.14%	**0.5%**	1.48%	0.68%	0.31%	2.15%
PER	μ	0.4%	0.05%	0.%	1.31%	0.%	0.%	3.79%	**91.41%**	0.%	0.%	0.1%
	σ	0.2%	0.15%	0.%	0.4%	0.%	0.%	1.96%	**0.23%**	0.%	0.%	0.3%
PIR	μ	0.%	0.%	4.19%	5.15%	0.%	0.15%	1.47%	0.%	**70.88%**	0.51%	3.9%
	σ	0.%	0.%	1.8%	2.28%	0.%	0.29%	1.47%	0.%	**1.51%**	0.34%	1.36%
STI	μ	11.94%	0.%	8.44%	0.11%	0.78%	1.56%	1.11%	0.%	0.06%	**42.11%**	5.67%
	σ	1.37%	0.%	7.47%	0.22%	1.56%	1.91%	0.43%	0.%	0.17%	**0.65%**	2.93%
VEJ	μ	14.17%	0.%	9.65%	5.73%	0.%	0.%	13.72%	0.%	2.33%	5.83%	**72.88%**
	σ	4.42%	0.%	2.37%	1.43%	0.%	0.%	2.13%	0.%	0.54%	1.73%	**0.67%**

Table C 9 The confusion matrix of the highest overall accuracy rate of rabbit grey CRF using SVM

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**57.77%**	0.%	4.59%	0.%	0.%	7.82%	4.15%	0.%	0.%	2.36%	23.32%
	σ	**1.51%**	0.%	0.35%	0.%	0.%	0.57%	0.63%	0.%	0.%	0.21%	1.06%
EXI	μ	0.%	**99.66%**	0.%	0.%	0.%	0.34%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.1%	0.%	**85.81%**	0.12%	0.%	10.29%	0.92%	0.%	0.%	1.65%	1.11%
	σ	0.16%	0.%	**0.61%**	0.12%	0.%	0.6%	0.21%	0.%	0.%	0.1%	0.12%
INT	μ	0.%	0.%	0.75%	**78.51%**	0.%	5.9%	2.76%	0.%	8.66%	0.%	3.43%
	σ	0.%	0.%	0.%	**1.15%**	0.%	0.85%	0.34%	0.%	0.68%	0.%	0.37%
IRR	μ	0.%	0.%	0.%	0.%	**64.4%**	35.6%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.%	**0.92%**	0.92%	0.%	0.%	0.%	0.%	0.%
MAG	μ	0.%	0.%	3.45%	0.%	0.36%	**95.85%**	0.%	0.%	0.%	0.34%	0.%
	σ	0.%	0.%	0.31%	0.%	0.%	**0.35%**	0.%	0.%	0.%	0.05%	0.%
MED	μ	2.4%	0.%	4.73%	0.34%	0.%	2.43%	**84.14%**	1.71%	0.%	0.14%	4.11%
	σ	0.34%	0.%	0.3%	0.%	0.%	0.36%	**0.8%**	0.%	0.%	0.27%	0.43%
PER	μ	0.%	0.%	0.%	0.%	0.%	3.33%	2.78%	**93.89%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.%	0.%	0.46%	0.25%	**0.66%**	0.%	0.%	0.%
PIR	μ	0.15%	0.%	5.29%	5.74%	0.%	4.49%	3.38%	0.%	**79.93%**	0.%	1.03%
	σ	0.29%	0.%	0.64%	0.79%	0.%	0.4%	0.36%	0.%	**0.87%**	0.%	0.36%
STI	μ	6.28%	0.%	10.28%	0.%	0.%	13.39%	0.%	0.%	0.56%	**63.61%**	5.89%
	σ	1.%	0.%	1.43%	0.%	0.%	1.88%	0.%	0.%	0.%	**1.83%**	0.87%
VEJ	μ	8.65%	0.%	2.26%	0.66%	0.%	1.53%	6.11%	0.%	0.03%	1.11%	**79.65%**
	σ	0.57%	0.%	0.5%	0.1%	0.%	0.17%	0.82%	0.%	0.1%	0.21%	**1.22%**

Table C 10 The confusion matrix of the highest overall accuracy rate of rabbit RGB CF using SVM

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**77.51%**	0.%	2.36%	0.48%	0.79%	0.61%	4.67%	0.%	0.09%	3.23%	10.26%
	σ	**4.93%**	0.%	0.88%	0.6%	0.51%	0.59%	3.19%	0.%	0.26%	1.%	1.97%
EXI	μ	0.%	**100.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.%**	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.75%	0.%	**92.2%**	0.29%	0.12%	2.13%	1.45%	0.%	0.27%	1.4%	1.38%
	σ	0.44%	0.%	**1.82%**	0.15%	0.29%	0.76%	1.16%	0.%	0.3%	0.84%	0.5%
INT	μ	0.15%	0.07%	0.%	**95.%**	0.%	0.%	0.82%	0.67%	2.39%	0.15%	0.75%
	σ	0.3%	0.22%	0.%	**2.31%**	0.%	0.%	0.97%	0.62%	2.05%	0.3%	0.82%
IRR	μ	0.43%	0.%	0.22%	0.%	**96.25%**	2.5%	0.%	0.%	0.%	0.22%	0.38%
	σ	0.33%	0.%	0.36%	0.%	**1.13%**	0.65%	0.%	0.%	0.%	0.27%	0.25%
MAG	μ	0.13%	0.%	2.37%	0.04%	1.04%	**94.96%**	0.13%	0.%	0.%	0.77%	0.56%
	σ	0.14%	0.%	0.65%	0.07%	0.38%	**1.31%**	0.12%	0.%	0.%	0.5%	0.56%
MED	μ	2.6%	0.07%	2.26%	0.21%	0.%	0.14%	**89.49%**	1.99%	0.31%	0.55%	2.4%
	σ	1.11%	0.14%	1.28%	0.41%	0.%	0.23%	**2.26%**	0.59%	0.32%	0.53%	1.22%
PER	μ	0.%	0.%	0.%	0.05%	0.%	0.%	1.87%	**98.08%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.15%	0.%	0.%	1.28%	**1.35%**	0.%	0.%	0.%
PIR	μ	0.37%	0.%	0.44%	4.49%	0.%	0.%	0.51%	0.07%	**92.5%**	0.29%	1.32%
	σ	0.37%	0.%	0.94%	2.14%	0.%	0.%	0.47%	0.22%	**2.18%**	0.36%	0.55%
STI	μ	3.72%	0.%	3.67%	0.11%	0.22%	3.22%	1.11%	0.%	0.06%	**87.11%**	0.78%
	σ	2.24%	0.%	2.05%	0.22%	0.44%	1.7%	1.47%	0.%	0.17%	**4.17%**	0.67%
VEJ	μ	5.%	0.%	1.22%	0.8%	0.%	0.1%	2.74%	0.%	0.14%	0.17%	**89.83%**
	σ	1.2%	0.%	0.64%	0.52%	0.%	0.22%	1.06%	0.%	0.23%	0.32%	**1.72%**

Table C 11 The confusion matrix of the highest overall accuracy rate of rabbit RGB RF using SVM

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**65.76%**	0.09%	3.06%	2.1%	1.14%	0.17%	10.66%	0.%	0.26%	4.8%	11.97%
	σ	**5.65%**	0.26%	0.85%	1.45%	0.45%	0.35%	3.38%	0.%	0.35%	1.44%	3.14%
EXI	μ	0.%	**99.97%**	0.%	0.03%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.1%**	0.%	0.1%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.27%	0.%	**90.58%**	0.34%	0.31%	4.24%	0.77%	0.02%	1.14%	0.87%	1.45%
	σ	0.23%	0.%	**2.54%**	0.22%	0.27%	1.5%	1.15%	0.07%	0.68%	0.6%	0.63%
INT	μ	0.82%	0.22%	0.%	**89.55%**	0.%	0.%	3.88%	0.75%	2.09%	0.97%	1.72%
	σ	0.78%	0.34%	0.%	**2.14%**	0.%	0.%	1.94%	0.75%	1.85%	0.89%	0.82%
IRR	μ	0.43%	0.%	1.25%	0.%	**94.95%**	2.55%	0.%	0.%	0.%	0.54%	0.27%
	σ	0.41%	0.%	1.%	0.%	**1.24%**	1.%	0.%	0.%	0.%	0.54%	0.27%
MAG	μ	0.04%	0.%	3.42%	0.%	1.4%	**94.55%**	0.02%	0.%	0.%	0.4%	0.18%
	σ	0.07%	0.%	0.75%	0.%	0.38%	**1.29%**	0.05%	0.%	0.%	0.49%	0.43%
MED	μ	4.21%	0.07%	1.06%	1.13%	0.%	0.07%	**86.2%**	2.12%	0.17%	1.95%	3.01%
	σ	1.41%	0.14%	1.49%	0.74%	0.%	0.14%	**2.49%**	0.76%	0.23%	1.05%	0.59%
PER	μ	0.%	0.%	0.%	0.05%	0.%	0.%	3.13%	**96.82%**	0.%	0.%	0.%
	σ	0.%	0.%	0.%	0.15%	0.%	0.%	1.49%	**1.48%**	0.%	0.%	0.%
PIR	μ	0.22%	0.07%	2.57%	2.28%	0.%	0.%	0.51%	0.07%	**92.57%**	0.07%	1.62%
	σ	0.34%	0.22%	2.01%	2.78%	0.%	0.%	0.57%	0.22%	**4.43%**	0.22%	1.03%
STI	μ	4.72%	0.%	3.56%	0.44%	0.56%	1.78%	2.94%	0.%	0.06%	**85.61%**	0.33%
	σ	1.61%	0.%	2.24%	0.22%	0.66%	1.77%	1.27%	0.%	0.17%	**4.58%**	0.57%
VEJ	μ	4.79%	0.%	0.97%	1.84%	0.03%	0.%	4.24%	0.%	0.52%	0.%	**87.6%**
	σ	1.54%	0.%	0.99%	1.22%	0.1%	0.%	1.66%	0.%	0.32%	0.%	**2.31%**

Table C 12 The confusion matrix of the overall accuracy of rabbit RGB CRF using SVM classifier

		COE	EXI	FLA	INT	IRR	MAG	MED	PER	PIR	STI	VEJ
COE	μ	**82.75%**	0.%	1.44%	0.17%	0.7%	0.13%	4.89%	0.%	0.26%	1.83%	7.82%
	σ	**2.28%**	0.%	1.05%	0.73%	1.01%	0.28%	1.73%	0.%	0.4%	0.78%	2.77%
EXI	μ	0.%	**99.97%**	0.%	0.03%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
	σ	0.%	**0.1%**	0.%	0.1%	0.%	0.%	0.%	0.%	0.%	0.%	0.%
FLA	μ	0.17%	0.%	**95.71%**	0.24%	0.02%	1.74%	0.44%	0.%	0.17%	0.7%	0.8%
	σ	0.33%	0.%	**1.4%**	0.11%	0.07%	0.79%	0.18%	0.%	0.16%	0.45%	0.31%
INT	μ	0.07%	0.%	0.%	**94.33%**	0.%	0.%	2.39%	0.07%	1.49%	0.%	1.64%
	σ	0.32%	0.%	0.%	**2.74%**	0.%	0.%	1.15%	0.22%	1.2%	0.%	1.33%
IRR	μ	0.16%	0.%	0.%	0.%	**96.96%**	2.88%	0.%	0.%	0.%	0.%	0.%
	σ	0.25%	0.%	0.%	0.%	**1.75%**	1.35%	0.%	0.%	0.%	0.%	0.%
MAG	μ	0.%	0.%	1.75%	0.02%	1.6%	**95.97%**	0.%	0.%	0.%	0.61%	0.05%
	σ	0.%	0.%	0.55%	0.05%	0.31%	**0.76%**	0.%	0.%	0.%	0.12%	0.05%
MED	μ	2.02%	0.%	0.34%	0.41%	0.%	0.%	**92.67%**	1.75%	0.1%	0.51%	2.19%
	σ	0.68%	0.%	0.27%	0.21%	0.%	0.%	**1.31%**	0.66%	0.16%	0.28%	0.71%
PER	μ	0.%	2.17%	0.%	0.%	0.%	0.%	4.8%	**93.03%**	0.%	0.%	0.%
	σ	0.%	0.15%	0.%	0.%	0.%	0.%	1.17%	**1.15%**	0.%	0.%	0.%
PIR	μ	0.%	0.%	3.46%	3.24%	0.%	0.07%	0.88%	0.%	**90.59%**	0.%	1.76%
	σ	0.%	0.%	0.44%	1.%	0.%	0.22%	0.36%	0.%	**1.46%**	0.%	0.75%
STI	μ	3.44%	0.%	2.17%	0.%	0.72%	1.33%	0.89%	0.%	0.%	**91.22%**	0.22%
	σ	2.39%	0.%	1.14%	0.%	0.44%	0.94%	0.61%	0.%	0.%	**4.54%**	0.27%
VEJ	μ	0.95%	0.%	1.11%	0.31%	0.%	0.%	3.58%	0.%	0.49%	3.42%	**90.14%**
	σ	0.86%	0.%	0.33%	0.38%	0.%	0.%	0.69%	0.%	0.22%	0.23%	**2.5%**

Appendix D: Classification results obtained through the minimisation of the images

Table D 1 Overall accuracy rates for classification chicken image datasets when Bilinear interpolation was used to minimize image sizes (Chicken feature sets)

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (340×247)	K-NN	μ	84.29%	79.3%	85.53%	94.55%	90.13%	94.47%
		σ	±1.56%	±1.56%	±2.33%	±2.33%	±2.01%	±2.16%
	ANN	μ	87.36%	79.03%	88.77%	92.33%	88.05%	93.98%
		σ	±1.64%	±1.93%	±2.25%	±1.47%	±1.97%	±0.78%
	SVM	μ	82.68%	81.37%	85.75%	93.28%	90.3%	94.08%
		σ	±1.82%	±2.03%	±1.98%	±2.39%	±1.72%	±2.48%
½ (255×185)	K-NN	μ	82.8%	79.97%	84.03%	94.08%	90.77%	94.17%
		σ	±1.54%	±2.03%	±2.66%	±2.27%	±2.14%	±2.07%
	ANN	μ	86.63%	79.25%	88.85%	92.75%	88.32%	94.07%
		σ	±1.59%	±2.08%	±2.52%	±1.62%	±1.4%	±0.81%
	SVM	μ	81.3%	81.21%	84.88%	93.24%	91.27%	94.14%
		σ	±1.75%	±1.75%	±2.14%	±2.47%	±1.82%	±2.46%
⅓ (170×123)	K-NN	μ	83.34%	79.99%	84.03%	93.04%	90.07%	94.32%
		σ	±1.85%	±1.86%	±2.63%	±2.07%	±2.37%	±2.29%
	ANN	μ	85.67%	79.02%	87.93%	92.49%	88.18%	93.92%
		σ	±2.0%	±1.89%	±2.29%	±1.71%	±1.78%	±1.03%
	SVM	μ	81.99%	81.6%	84.39%	92.56%	90.87%	92.88%
		σ	±2.09%	±1.93%	±2.26%	±2.89%	±1.57%	±1.9%
¼ (128×93)	K-NN	μ	82.53%	79.99%	83.82%	93.72%	90.09%	93.41%
		σ	±1.79%	±1.86%	±2.38%	±2.42%	±2.52%	±2.11%
	ANN	μ	88.08%	79.53%	88.53%	92.69%	88.04%	94.27%
		σ	±1.27%	±1.46%	±2.27%	±1.9%	±2.02%	±0.77%
	SVM	μ	82.71%	80.77%	84.06%	93.6%	90.48%	93.86%
		σ	±1.77%	±2.06%	±1.82%	±2.31%	±2.04%	±2.26%
1/5 (102×74)	K-NN	μ	63.2%	60.68%	64.08%	64.69%	62.79%	66.21%
		σ	±3.49%	±3.54%	±3.47%	±3.49%	±3.52%	±3.63%
	ANN	μ	58.93%	57.96%	62.53%	65.48%	60.4%	67.3%
		σ	±2.92%	±2.52%	±2.66%	±3.08%	±3.39%	±2.93%
	SVM	μ	59.27%	58.05%	60.51%	61.27%	59.24%	62.26%
		σ	±2.77%	±2.41%	±2.9%	±2.7%	±2.87%	±2.53%
1/10 51×37	K-NN	μ	20.45%	16.98%	21.4%	21.84%	19.68%	23.81%
		σ	±4.7%	±4.98%	±4.83%	±5.06%	±5.01%	±4.99%
	ANN	μ	20.11%	19.47%	20.98%	21.99%	19.61%	23.39%
		σ	±4.7%	±4.61%	±4.63%	±4.73%	±4.55%	±4.78%
	SVM	μ	19.7%	16.9%	21.6%	21.2%	20%	23.02%
		σ	±3.94%	±3.96%	±4.09%	±4.01%	±4.08%	±3.96%

Table D 2 Overall accuracy rates for classification chicken image datasets when Bicubic interpolation was used to minimize image sizes (Chicken feature sets)

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (340×247)	K-NN	μ	85.13%	79.91%	85.15%	94.2%	91.12%	94.73%
		σ	±2.58%	±1.78%	±2.37%	±2.25%	±2.17%	±1.97%
	ANN	μ	87.31%	79.21%	88.81%	91.68%	88.17%	94.06%
		σ	±1.91%	±1.95%	±2.25%	±2.17%	±2.59%	±1.28%
	SVM	μ	81.84%	81.02%	85.%	94.6%	91.29%	94.16%
		σ	±1.96%	±1.84%	±1.99%	±2.48%	±1.96%	±2.27%
½ (255×185)	K-NN	μ	82.48%	79.25%	84.26%	94.11%	90.51%	94.66%
		σ	±1.78%	±1.37%	±2.91%	±2.02%	±1.81%	±2.12%
	ANN	μ	87.08%	79.39%	88.93%	92.4%	89.28%	94.%
		σ	±1.59%	±2.04%	±2.39%	±1.98%	±1.96%	±1.51%
	SVM	μ	81.41%	80.61%	84.72%	93.05%	91.06%	93.55%
		σ	±2.04%	±2.08%	±2.12%	±2.37%	±1.97%	±2.38%
⅓ (170×123)	K-NN	μ	83.97%	79.88%	84.63%	94.03%	90.57%	94.1%
		σ	±2.19%	±1.95%	±2.9%	±2.38%	±2.33%	±1.73%
	ANN	μ	87.36%	79.43%	87.46%	92.42%	87.18%	93.31%
		σ	±1.66%	±1.88%	±2.65%	±2.09%	±2.39%	±1.81%
	SVM	μ	81.85%	81.48%	84.73%	93.34%	90.52%	93.75%
		σ	±1.73%	±1.69%	±2.21%	±2.38%	±2.05%	±2.26%
¼ (128×93)	K-NN	μ	83.03%	79.32%	84.56%	93.92%	90.81%	93.77%
		σ	±2.15%	±1.44%	±2.72%	±2.16%	±1.98%	±2.45%
	ANN	μ	88.23%	79.71%	88.75%	92.51%	88.18%	93.27%
		σ	±2.13%	±2.03%	±2.22%	±2.03%	±2.14%	±1.57%
	SVM	μ	81.85%	79.97%	84.66%	92.89%	90.83%	93.25%
		σ	±1.97%	±1.86%	±2.25%	±2.7%	±1.81%	±2.53%
$1/5$ (102×74)	K-NN	μ	59.18%	59.4%	61.0%	62.84%	60.79%	63.96%
		σ	±3.91%	±3.91%	±3.93%	±3.94%	±4.08%	±3.88%
	ANN	μ	60.04%	58.1%	62.01%	65.8%	59.08%	65.87%
		σ	±3.36%	±4.29%	±3.8%	±4.0%	±4.6%	±3.64%
	SVM	μ	60.02%	57.87%	60.3%	60.18%	57.95%	61.0%
		σ	±3.13 %	±3. 68%	±3.34%	±3.66%	±3.18%	±3.55%
$1/10$ (51×37)	K-NN	μ	16.49%	15.3%	19.4%	19.3%	15.3%	19.77%
		σ	±4.82%	±5.1%	±4.89%	±5.09%	±4.83%	±4.88%
	ANN	μ	19.98%	18.6%	20.49%	21.97%	19.0%	24.3%
		σ	±4.08%	±4.45%	±3.9%	±4.04%	±4.07%	±3.96%
	SVM	μ	19.9%	18.3%	20.9%	21.66%	20.1%	22.76%
		σ	±3.92%	±3.98%	±3.96%	±3.94%	±3.96%	±3.95%

Table D 3 Overall accuracy rates for classification chicken image datasets when Nearest Neighbour interpolation was used to minimize image sizes (Chicken feature sets)

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ (321 × 223)	K-NN	μ	69.64%	63.39%	79.87%	83.11%	78.33%	82.84%
		σ	±0.67%	±1.55%	±2.1%	±2.16%	±5.39%	±2.25%
	ANN	μ	76.34%	71.33%	82.73%	91.19%	89.86%	93.05%
		σ	±4.01%	±2.31%	±5.07%	±3.36%	±4.78%	±1.75%
	SVM	μ	77.62%	71.08%	82.9%	92.02%	85.6%	92.49%
		σ	±1.02%	±1.23%	±2.29%	±1.99%	±2.6%	±2.15%
½ (241 × 168)	K-NN	μ	69.87%	63.36%	79.47%	82.54%	77.13%	82.92%
		σ	±0.61%	±1.23%	±2.34%	±2.46%	±5.66%	±2.49%
	ANN	μ	76.61%	71.41%	80.56%	91.26%	88.64%	93.98%
		σ	±3.09%	±3.21%	±5.01%	±3.13%	±4.01%	±2.46%
	SVM	μ	77.65%	71.07%	82.66%	91.84%	85.11%	92.82%
		σ	±0.93%	±1.54%	±2.07%	±1.82%	±2.69%	±1.85%
⅓ (161 × 112)	K-NN	μ	69.91%	63.49%	79.28%	82.96%	77.89%	82.62%
		σ	±0.63%	±1.63%	±2.35%	±2.75%	±5.24%	±2.23%
	ANN	μ	74.67%	70.73%	80.54%	90.36%	88.31%	93.42%
		σ	±3.47%	±3.25%	±5.02%	±2.81%	±4.82%	±2.84%
	SVM	μ	77.7%	70.99%	82.65%	92.25%	85.77%	92.65%
		σ	±0.97%	±1.61%	±2.26%	±2.32%	±3.07%	±1.72%
¼ (121 × 84)	K-NN	μ	69.94%	62.71%	79.34%	82.74%	77.72%	82.71%
		σ	±0.64%	±1.42%	±2.09%	±2.6%	±5.67%	±2.82%
	ANN	μ	74.8%	70.06%	80.23%	90.0%	88.95%	92.88%
		σ	±3.78%	±2.11%	±4.66%	±3.29%	±5.48%	±2.05%
	SVM	μ	77.63%	70.74%	82.43%	91.99%	84.65%	92.82%
		σ	±0.83%	±1.72%	±2.44%	±2.11%	±2.72%	±1.86%
$1/5$ (96 × 67)	K-NN	μ	45.98%	43.20%	47.50%	49.30%	47.62%	52.46%
		σ	±3.56%	±4.4%	±3.51%	±3.24%	±4.25%	±4.47%
	ANN	μ	47.6%	45.3%	49.8%	51.73%	49.78%	53.5%
		σ	±4.52%	±4.20%	±4.60%	±4.9%	±4.82%	±4.93%
	SVM	μ	47.1%	45.8%	50.1%	51.74%	48.87%	54.19%
		σ	±2.77%	±2.83%	±2.54%	±2.89%	±2.9%	±3.12%
$1/10$ (48 × 34)	K-NN	μ	19.80%	16.20%	22.85%	22.40%	21.97%	25.60%
		σ	±6.26%	±7.1%	±6.12%	±6.25%	±6.87%	±6.2%
	ANN	μ	22.5%	20.8%	23.58%	23.8%	21.2%	27.7%
		σ	±5.85%	±6.0%	±5.6%	±6.25%	±6.82%	±7.25%
	SVM	μ	20.48%	18.25%	22.85%	23.5%	19.4%	26.4%
		σ	±3.25%	±3.89%	±3.65%	±3.20%	±3.85%	±3.56%

Table D 4 Overall accuracy rates for classification chicken image datasets when Bilinear Interpolation was used to minimize image sizes (Rabbit feature sets)

Minimizing Degree (Pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ 321 × 223	K-NN	μ	70.25%	64.84%	79.57%	83.32%	76.39%	83.81%
		σ	±2.73%	±2.08%	±1.8%	±2.89%	±6.08%	±3.13%
	ANN	μ	77.06%	71.13%	83.06%	92.63%	89.52%	94.77%
		σ	±2.82%	±2.78%	±5.44%	±2.71%	±4.3%	±1.93%
	SVM	μ	77.7%	71.26%	82.89%	92.84%	85.47%	92.66%
		σ	±0.86%	±1.25%	±2.48%	±1.79%	±2.87%	±1.69%
½ 241 × 168	K-NN	μ	69.82%	64.03%	79.36%	83.32%	77.45%	83.17%
		σ	±2.63%	±2.02%	±2.18%	±2.89%	±5.68%	±3.17%
	ANN	μ	76.14%	71.76%	81.3%	91.99%	89.41%	93.96%
		σ	±3.36%	±2.89%	±5.1%	±3.27%	±5.03%	±2.05%
	SVM	μ	77.89%	70.18%	82.4%	92.33%	85.91%	92.29%
		σ	±1.01%	±1.55%	±2.59%	±2.01%	±2.72%	±1.59%
⅓ 161 × 112	K-NN	μ	69.32%	63.77%	78.98%	83.32%	75.93%	82.31%
		σ	±2.55%	±1.91%	±1.94%	±2.89%	±7.2%	±2.66%
	ANN	μ	75.61%	71.49%	80.63%	91.29%	87.37%	93.45%
		σ	±3.04%	±2.51%	±4.61%	±3.84%	±4.6%	±1.97%
	SVM	μ	77.92%	70.87%	81.89%	92.16%	84.56%	92.97%
		σ	±0.78%	±1.05%	±2.32%	±1.75%	±3.04%	±2.02%
¼ 121 × 84	K-NN	μ	69.05%	62.94%	79.73%	81.84%	74.54%	81.61%
		σ	±2.88%	±2.1%	±2.08%	±2.09%	±4.77%	±3.18%
	ANN	μ	75.44%	70.12%	80.84%	90.23%	87.32%	93.69%
		σ	±2.88%	±3.25%	±5.13%	±3.03%	±4.04%	±1.85%
	SVM	μ	77.96%	70.85%	81.63%	91.83%	85.59%	92.43%
		σ	±0.95%	±1.42%	±2.01%	±1.74%	±2.67%	±1.88%
$1/5$ 96 × 67	K-NN	μ	45.98%	44.3%	47.76%	49.3%	46.28%	51.67%
		σ	±3.11%	±3.8%	±3.4%	±3.18%	±3.87%	±4.11%
	ANN	μ	47.51%	45.25%	49.44%	51.24%	49.08%	53.50%
		σ	±4.20%	±3.89%	±4.0%	±3.76%	±4.1%	±4.83%
	SVM	μ	47.5%	45.2%	49.4%	52%	48.2%	55.6%
		σ	±2.18%	±3.2%	±2.43%	±2.22%	±3.1%	±3.4%
$1/10$ 48 × 34	K-NN	μ	20.02%	17.08%	21.45%	21.95%	21.05%	25.56%
		σ	±5.94%	±6.73%	±6.23%	±5.8%	±6.23%	±5.92%
	ANN	μ	21.93%	19.94%	23.59%	23.71%	22.08%	26.86%
		σ	±5.70%	±5.98%	±5.11%	±5.87%	±6.3%	±6.92%
	SVM	μ	21.13%	18.79%	23.24%	23.35%	20.83%	26.37%
		σ	±3.55%	±3.62%	±3.18%	±3.56%	±3.47%	±3.22%

Table D 5 Overall accuracy rates for classification chicken image datasets when Bicubic interpolation was used to minimize image sizes (Rabbit feature sets)

Minimizing Degree (pixels)	Classifier		Feature Sets					
			Grey CF	Grey RF	Grey CRF	RGB CF	RGB RF	RGB CRF
⅔ 321 × 223	K-NN	μ	70.55%	64.84%	79.85%	82.84%	76.26%	83.98%
		σ	±2.22%	±2.08%	±1.64%	±2.01%	±7.1%	±3.41%
	ANN	μ	76.34%	71.32%	82.69%	91.75%	88.76%	93.96%
		σ	±3.52%	±2.61%	±3.64%	±3.15%	±5.31%	±2.36%
	SVM	μ	77.59%	71.46%	82.43%	92.36%	85.71%	92.83%
		σ	±1.05%	±1.5%	±2.88%	±1.97%	±2.85%	±1.81%
½ 241 × 168	K-NN	μ	69.53%	64.03%	79.55%	82.42%	76.52%	82.14%
		σ	±2.27%	±2.02%	±2.08%	±1.64%	±6.7%	±2.05%
	ANN	μ	74.77%	70.94%	82.69%	90.92%	88.43%	93.58%
		σ	±3.39%	±2.88%	±3.76%	±3.9%	±5.3%	±2.33%
	SVM	μ	77.9%	70.78%	82.65%	92.43%	85.2%	92.91%
		σ	±1.07%	±1.42%	±2.8%	±1.37%	±2.79%	±2.05%
⅓ 161 × 112	K-NN	μ	69.74%	63.77%	79.51%	82.73%	75.49%	83.32%
		σ	±2.37%	±1.91%	±2.21%	±1.92%	±5.95%	±2.89%
	ANN	μ	75.06%	70.63%	81.64%	89.82%	87.93%	93.72%
		σ	±3.78%	±3.3%	±3.31%	±3.44%	±4.81%	±1.62%
	SVM	μ	77.99%	70.31%	81.89%	91.91%	85.5%	92.89%
		σ	±0.88%	±1.62%	±2.32%	±1.97%	±2.51%	±1.84%
¼ 121 × 84	K-NN	μ	69.45%	62.94%	79.79%	82.66%	76.35%	82.38%
		σ	±2.51%	±2.1%	±2.0%	±1.99%	±5.75%	±2.96%
	ANN	μ	74.81%	70.75%	81.57%	90.57%	88.36%	93.52%
		σ	±3.09%	±3.07%	±3.89%	±3.89%	±5.33%	±2.12%
	SVM	μ	77.91%	70.7%	82.19%	91.44%	84.78%	92.45%
		σ	±0.83%	±1.37%	±2.29%	±2.0%	±2.66%	±1.89%
$1/5$ 96 × 67	K-NN	μ	45.98%	43.75%	47.63%	49.3%	46.95%	52.07%
		σ	±3.37%	±4.13%	±3.76%	±3.19%	±4.17%	±4.29%
	ANN	μ	46.98%	45.62%	48.98%	51.80%	49.76%	54.10%
		σ	±3.99%	±4.02%	±4.04%	±4.15%	±4.54%	±4.22%
	SVM	μ	47.19%	45.54%	49.90%	51.84%	48.94%	54.62%
		σ	±3.57%	±3.65%	±3.28%	±3.04%	±3.74%	±3.09%
$1/10$ 48 × 34	K-NN	μ	19.91%	16.64%	22.15%	22.18%	21.51%	25.58%
		σ	±4.29%	±5.06%	±4.44%	±4.04%	±4.49%	±4.78%
	ANN	μ	21.0%	18.20%	23.30%	22.84%	21.01%	25.85%
		σ	±5.60%	±6.02%	±5.48%	±5.62%	±6.25%	±5.70%
	SVM	μ	20.87%	18.41%	23.13%	23.23%	20.41%	26.21%
		σ	±3.67%	±4.08%	±3.17%	±3.55%	±4.00%	±3.97%

Appendix E: The positions of the super-pixels selected using Relieff feature selection method

Figure E 1 Samples of the Eimeria species of chickens with different number of the most relevant super-pixels for the original size images